

77634

10/4 04 + 02

Access DB#

10/4 115

## SEARCH REQUEST FORM

Scientific and Technical Information Center

78201

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: Oct 4, 2002  
 Art Unit: 1652 Phone Number 305-7270 Serial Number: 09/720,583  
 Mail Box and Bldg/Room Location: 10 D.O.G Results Format Preferred (circle): PAPER DISK E-MAIL  
10.D01

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Propionibacterium Vector  
 Inventors (please provide full names): Pieter Pauwels et al

Earliest Priority Filing Date: 6/25/99

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO: 1 DNA  
 2 AA  
 3 AA

Thank you so much in advance.

*M. Walicka*

Point of Contact:  
 Toby Port  
 Technical Info. Specialist  
 CM1 6A04  
 703-308-3634

## STAFF USE ONLY

## Type of Search

## Vendors and cost where applicable

Searcher: _____	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>2</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>10/4</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>10/7</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>12</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>12</u>	Other _____	Other (specify) _____

**THIS PAGE BLANK (USPTO)**



```

/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07380.1"
/db_xref="GI:9995030"
/translation="MSQFETLPFESWLPKRLPLASAEKSGAVRHVTRORALELPEYDAN
PLWQSLVIFTRDSDADMAADLADGLSPSYSNRRYTTGHIYYALKNPVCLTDAR
RPRNPLNARVEOGICDVLGGDASGHHITLSTATATLNGPADALYELRALAHITD
EHALPEAGNPRPRNRVPSYGRVNTLEPTDRMAYRVARHSMGCPVAEMHHYDPEHHT
LNNETITADEPFGNPGDLNKLKHSISISWVNRNFTPEFRANOKAISLNGASKGK
EGGKGGIAGSAGRAHTRQOFLEGLS"
1181..1438
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07381.1"
/db_xref="GI:9995031"
/translation="MTTRERLPPRNGYSIAAANKLGVSESTVKRWISERPEEFVARYA
ARRAHIRELSESGSMALIAEAEGSVGTVAHLNKNRTDA"

```

Query Match	100.0%;	Score 3555;	DB 6;	Length 3555;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3555;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	
QY 1	gtgacacctgacacgcgagacgacgttcacggacaaagtgcacaagctgacgaggaacta	60		
Db 1	GTCAACCTGACACCCGGCGAGCATTTCCAGGCGAAGATCCACAGCTGCCCCGAGAACTA	60		
QY 61	gcgcgaatgccgacacacgcgcccgacatcccttgagacaggtgagcagctcaagggagt	120		
Db 61	GCGCGAATGCCCGAACAACGCCCGACCATCCCTTGAGACAGGTGGCAGCGCTCAGGGAGT	120		
QY 121	cggaggatgttttgcgacggagatgtgtgaagaagatgttcgtcttcacatgctcaacg	180		
Db 121	CGGGGAGTGTTTGGCAGGGGAGTGTGGAAAGAGATTGCTTTGGTCCATGAGGTCAACG	180		
QY 181	ggttaactcaactgtataatggtgtcttcgtcgcgccactttgaaacagccgaggaatgacac	240		
Db 181	GGTAACTCAACGAATATGGGGCTTCGTGCGCCACTTTGAACAGCGCGAATGAGCAC	240		
QY 241	gcgtaacgttgaactgcacgtcttcaactcatgtatgattgatttgcgagagctttccgt	300		
Db 241	GCTAACGTTGACTGCAATGCTTCACTCATATGATAGATTTGTTGAAACCTTTTCCCTG	300		
QY 301	agaagctgtctgcacacgaacgcgcgtctgcgcacgacggagaagctctgaggctacgcgacg	360		
Db 301	AGAACTGTGCTCCACGCAACGCCGTGGCGTCAAGCCGAAAGTGTGGGGCTTACCGGCAG	360		
QY 361	tgaactgcgcagagggcgctgagacgtgctctacatcgaaagcgaacccgttgtcatgagt	420		
Db 361	TGACTCGGCAAGGGCGCTGGAGCTGCCTTACATCGAAGCAACCGTGTGTCATGAGT	420		
QY 421	ccttggtcatalcaacgatacgagatgcttcgagatgtcagctgagcgcgacagctcgtgggc	480		
Db 421	CCTTGGTCAATCACGATTCGGAATGCTTCGATGCTGACTGGCGCGCAACTCGCTGGGC	480		
QY 481	tgccttcaacgcttcctagctgttccatctgaacgcttcaagagcaacacgaactcgtctatg	540		
Db 481	TGCTTCAACGCTCTTACGTGTTCATGTAACCGTGTGCAGACACCGGCAACATGCTCATG	540		
QY 541	ccttgaagaacccgtgtgtctgcacgaatgtccgcgcgagcgagcgctatcaactcgtcg	600		
Db 541	CCTTGAAGAACCCTGTGTGTCTGATGCAATGCGCGCGGCAACGGCCCTATAACTGCTG	600		
QY 601	ccgcgcgtcagacagagggcctatgcagcttctcggcgcgatgatacctcaagcgacagga	660		
Db 601	CCCGCTCGAAGCAGGGCCTATGACAGCTCTCGGGCGCGATGATGATCCTACGGGCAACGG	660		
QY 661	tcacaaagaacccgctcagacacgcccatagtcgaacctcttgaggcccgacagcgcgtct	720		
Db 661	tcacaaagaacccgctcagacacgcccatagtcgaacctcttgaggcccgacagcgcgtct	720		

[illegible]



QY 1801 ccgagcgacgtagaacccgctccggttgggtggtgacacgctgacacgagtcgctt 1860  
|||  
Db 1801 ccgagcgacgtagaacccgctccggttgggtggtgacacgctgacacgagtcgctt 1860  
|||  
QY 1861 ttgggtgtgtagtagaacccctcgaatgccccggtgcatgtagcgttagttagtgagc 1920  
|||  
Db 1861 ttgggtgtgtagtagaacctcgaatgccccggtgcatgtagcgttagttagtgagc 1920  
|||  
QY 1921 tgcctctgagaaagagcttttaacggttcaaggtcagcagcagcagcagcagcagc 1980  
|||  
Db 1921 tgcctctgagaaagagcttttaacggttcaaggtcagcagcagcagcagcagcagc 1980  
|||  
QY 1981 ggcaggtgtagaacgacggttgcacagcagcagcagcagcagcagcagcagcagc 2040  
|||  
Db 1981 ggcaggtgtagaacgacggttgcacagcagcagcagcagcagcagcagcagcagc 2040  
|||  
QY 2041 ccgagcgacgtagaacgacgacgacgacgacgacgacgacgacgacgacgacgacg 2100  
|||  
Db 2041 ccgagcgacgtagaacgacgacgacgacgacgacgacgacgacgacgacgacgacg 2100  
|||  
QY 2101 ggtgagcagcgtcgttagcgttagcgttagcgttagcgttagcgttagcgttagc 2160  
|||  
Db 2101 ggtgagcagcgtcgttagcgttagcgttagcgttagcgttagcgttagcgttagc 2160  
|||  
QY 2161 aacaggtgctgacacagctgagcagcagcagcagcagcagcagcagcagcagc 2220  
|||  
Db 2161 aacaggtgctgacacagctgagcagcagcagcagcagcagcagcagcagcagc 2220  
|||  
QY 2221 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2280  
|||  
Db 2221 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2280  
|||  
QY 2281 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2340  
|||  
Db 2281 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2340  
|||  
QY 2341 actgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400  
|||  
Db 2341 actgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400  
|||  
QY 2401 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2460  
|||  
Db 2401 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2460  
|||  
QY 2461 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2520  
|||  
Db 2461 agtgcgtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2520  
|||  
QY 2521 gctcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2580  
|||  
Db 2521 gctcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2580  
|||  
QY 2581 ccgagcgacgtagaacgagcagcagcagcagcagcagcagcagcagcagcagc 2640  
|||  
Db 2581 ccgagcgacgtagaacgagcagcagcagcagcagcagcagcagcagcagcagc 2640  
|||  
QY 2641 gtagagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2700  
|||  
Db 2641 gtagagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2700  
|||  
QY 2701 ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2760  
|||  
Db 2701 ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2760  
|||  
QY 2761 cgtcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2820  
|||  
Db 2761 cgtcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2820  
|||  
QY 2821 aagttttcgtatagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2880  
|||  
Db 2821 aagttttcgtatagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2880  
|||

QY 2881 caacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2940  
|||  
Db 2881 caacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2940  
|||  
QY 2941 ccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000  
|||  
Db 2941 ccccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000  
|||  
QY 3001 ggtcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3060  
|||  
Db 3001 ggtcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3060  
|||  
QY 3061 gaaacgttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3120  
|||  
Db 3061 gaaacgttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3120  
|||  
QY 3121 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3180  
|||  
Db 3121 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3180  
|||  
QY 3181 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3240  
|||  
Db 3181 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3240  
|||  
QY 3241 tcatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3300  
|||  
Db 3241 tcatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3300  
|||  
QY 3301 acggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3360  
|||  
Db 3301 acggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3360  
|||  
QY 3361 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3420  
|||  
Db 3361 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3420  
|||  
QY 3421 atcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3480  
|||  
Db 3421 atcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3480  
|||  
QY 3481 gatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3540  
|||  
Db 3481 gatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3540  
|||  
QY 3541 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3540  
|||  
Db 3541 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3540  
|||

RESULT 2  
AX008203 3555 bp DNA linear PAT 06-SEP-2000  
LOCUS  
DEFINITION Sequence 1 from Patent WO967356.  
ACCESSION AX008203  
VERSION AX008203.1 GI:9995826  
KEYWORDS  
SOURCE  
ORGANISM  
Propionibacterium freudenreichii.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;  
Propionibacterium.  
REFERENCE  
JORE, J. P., LUTTEN, R. G., POWELS, P. H. and Van LUIJCK, N.  
1 (bases 1 to 3555)  
AUTHORS  
JORE, J. P., LUTTEN, R. G., POWELS, P. H. and Van LUIJCK, N.  
TITLE  
JOURNAL  
PATENT: WO 967356-A 1 29-DEC-1999;  
GIST BROCADES BV (NL); JORE JOHANNES PETRUS MARIA (NL); LUTIK  
NICOLE VAN (NL); LUTTEN RODOLF GUSBERTUS MARIE (NL); POWELS  
PIETER HENDRIK (NL)  
FEATURES  
location/Qualifiers  
1..3555  
/organism="Propionibacterium freudenreichii"  
/db\_xref="taxon:1744"  
/note="IMG16545"

```
CDS
273..1184
/note="unnamed protein product"
/codon_start=1
/cranal_table=11
/protein_id="CAC07396.1"
/db_xref="GI:9995827"
/translation="MDSFETLPPEEWLPKPLASAKSGAYRHVTRQRLALEPYEAN
PLVMSLVITDDASDADMAADLAGLPSPSYSMNKRVTTGHTVALKNPVCLTDAR
RRIINLAVNEQGLCDVLGDAISYGRITKNPLSTAHAILMGFADLALYELRALHTLD
EIHAIPEANPRNRYSTYGRNVLTFTDRMAAYAVHSMGSPVAWEHETVEFHH
LNETIADPEFATPLGLNELKLSISIRWWRNPETPTFAROKAISLRASKSGK
EGHKGGTASGASRAHTRQQLFEGLS"
1181..1438
/note="unnamed protein product"
/codon_start=1
/cranal_table=11
/protein_id="CAC07397.1"
/db_xref="GI:9995828"
/translation="MTTERLPRNGYSIAAAKKIGVSESTYKRWTSREPEFVARVA
ARRHRIRELRSBQSMRAIAAEVGSVGVHYALNNRIDA"
BASE COUNT      705 a 1156 c 1075 g 619 t
ORIGIN
Query Match      100.0%; Score 3555; DB 6; Length 3555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtcgaccctgacagccgagagcaagtcagcgagaaatgcacagctgcgagaggaacta 60
Db 1 gtcgacctgacagccgagagcaagtcagcgagaaatgcacagctgcgagaggaacta 60
QY 61 gccgcaatgcccgaacacagcccccagccatcccttgagacagtgctcagcgctcagggagat 120
Db 61 gccgcaatgcccgaacacagcccccagccatcccttgagacagtgctcagcgctcagggagat 120
QY 121 cggggagatgtttgacagggatgtgaaagagatgcgcttgctcacaatgctcaacg 180
Db 121 cggggagatgtttgacagggatgtgaaagagatgcgcttgctcacaatgctcaacg 180
QY 181 gtaactaactgatatgaggtcttcgtcgcccaacttggaaacagccgagagaaatggaaccac 240
Db 181 gtaactaactgatatgaggtcttcgtcgcccaacttggaaacagccgagagaaatggaaccac 240
QY 241 gctgaacgttgactcgcaatgcttcacatgtaataatgcttcgctcgaagcgtgttcctctg 300
Db 241 gctgaacgttgactcgcaatgcttcacatgtaataatgcttcgctcgaagcgtgttcctctg 300
QY 301 agagcgtgctgcacagcaagccgctgctcagcgagaaagctctggggctgaacgacg 360
Db 301 agagcgtgctgcacagcaagccgctgctcagcgagaaagctctggggctgaacgacg 360
QY 361 tgaactcgagcagagggcgctgagctgacctacatcgaaagcaacccgttgatcatgact 420
Db 361 tgaactcgagcagagggcgctgagctgacctacatcgaaagcaacccgttgatcatgact 420
QY 421 ccttgatcatalcacaatgcgaatgcttcgaatgctgaatgctgcgacagacactcgtgggc 480
Db 421 ccttgatcatalcacaatgcgaatgcttcgaatgctgaatgctgcgacagacactcgtgggc 480
QY 481 tgcctcaccgctcctacgtgtcctacatgaacgctgtcagcaacacggaacacatcgtctatg 540
Db 481 tgcctcaccgctcctacgtgtcctacatgaacgctgtcagcaacacggaacacatcgtctatg 540
QY 541 ccttgaagaacccctgtgtgtctgacagatgcgagcgagagccatataaactgctg 600
Db 541 ccttgaagaacccctgtgtgtctgacagatgcgagcgagagccatataaactgctg 600
QY 601 cccgcgttcgagcagggcctcctatgcaagcttctcgcgagcgaatgcatcctacggcagccgga 660
Db 601 cccgcgttcgagcagggcctcctatgcaagcttctcgcgagcgaatgcatcctacggcagccgga 660
QY 661 tcacaaagaacccgctcagcaacgcccacatgcaacccctctggtggcccccagacagcgctct 720
```

```
Db 661 TCACAAAGAACCCGCTCAGCAGCCGACCCATGCGACCCCTTGGGGCCCCCGCAGAGCCCTCT 720
QY 721 aagaagctgcgagccctcgcagacacacacctgcgaaggttccagcagctgcgagagcgaggga 780
Db 721 ACAAGCTGCCGCCCTCTCGCACACACCTTCGACGAGATCCAGCATGCGCGGAGGCGAGGGA 780
QY 781 acccgctgcgaacgtcaccggaatcaacaggttcggccgcaacgctcaacccgttctgaacaca 840
Db 781 ACCCGGTCGCAACGTACACCCGATCAACGGTGGCGCAACGTCACCCCTTCTTGACACCA 840
QY 841 cccgcatgtggaatacagcagcgctgcgacatcctcgtggggtggccggttcgccaatggg 900
Db 841 CCGCATGTGTGGCATACCGGGCGGTCCGGCATCTCGGGCGCGCCGCGGTTCGCGGAATGGG 900
QY 901 agaacaccgctattcggagacatccactactgaagagaaatcatalcgcgagaaattg 960
Db 901 AGCACACCGTATTGAGACATCCACTACTGAACGAGACGATCATGCGCGAGAAATTCG 960
QY 961 ccaagggccccctcggcttgaaacgaactaagcactatctcgatccatttcccgatggg 1020
Db 961 CCAAGGGCCCCCTCGGCTTGAAGCACTTAAAGCACTTATCGATCATTTCCGATGGG 1020
QY 1021 tctgagcaacttcaaccccgaaacccctccgagcagcagaaagcagatcagctcgtg 1080
Db 1021 TCTGAGCAACTTCAACCCCGAAACCTTCCGCGACGCGCAGAAAGCGATCAGCTCCGTG 1080
QY 1081 gaacatccaaagcgcgcaagaaagcgccacaaagggcgcatgtgcagtgagcagcagc 1140
Db 1081 GAGCATCCAAAGCGCGGGAAGAAAGCGCGCCACAAAGCGGCAATTCGAGTGGCGCATCAC 1140
QY 1141 ggcgcgcccaataccgctcaacagcttcttgaggggtctctcatatgacacacgctgaagctc 1200
Db 1141 GCGCGCGCCATACCCGTCGAACACGTTCTTGAGGGGTCTCTATACCAACAGTAAGTGT 1200
QY 1201 ccccgcgaacgctcaacgactcgcgcgtctgctgcgaaaagctcgtgtgctcgaatccac 1260
Db 1201 CCCCCGCAACGGGTACAGCATCCGCCCTGCTCGGAAAAACCTCGGTCTCCCAAGTCCAC 1260
QY 1261 cgtcaagcgtgtgacttcgagccagcagcgagaggttcgtggccgcggttcgcaagcgca 1320
Db 1261 GGTCAACGGGTGACATTCGAGCCACGCGAGAGGATTCGTGGCCCCGCTGGCCACGCCA 1320
QY 1321 cgcgcgattcgtgaagctccgctcgcgagaggttcagaacatgctgctgcagatgtc 1380
Db 1321 CCGCGGATTCGTGAAGCTCCGCTCGGAGGGTCAAGAGATCGTGGCATGTCCTCCGAGGT 1380
QY 1381 cggggtttccgtggtgcacgctgcactaagcgcgtgcgaagaatcgaactgaagcagatgacc 1440
Db 1381 CGGGGTTTCCGTGGGCAACCTGTGACATACCGCGTGAACAAAGATCGAACATGCATGACC 1440
QY 1441 gtaacgcgacagatgagcaatttcttgatcgtgcagccgcttggaactcagttcgtgctg 1500
Db 1441 GTAACGCCGACACATGAGCATTTTCTTGTGATCGTGACCCGCTTGGCATCTGCGGTGC 1500
QY 1501 ggttgacagtgctgcgcacagcttcttatacctgcgagcaatgtgtgctacagcacaatggggg 1560
Db 1501 GGTTCGACAGATGCGCGCACGCTTATCTCTGCGGCATGTGGGTACAGCCAAATGGGGG 1560
QY 1561 gcatcagcaaggaagcttgaaacccggttgggaaagtattactcaagggggagatgcccagtc 1620
Db 1561 GCATCAGCAAGGACGTTGAACCCGGTGGGCAAGTGTATCTAGGGGGGAGATGCCAGTTC 1620
QY 1621 tgcgcgcctcgaatgacagtatggagatgcgtgcagcgccccacgcgtcaaatcatttc 1680
Db 1621 TCGCGGCTCGGATTGACGGTATGCGAGTGTGCAATGCGGCCCCACCGTCAAACTATTTC 1680
QY 1681 aggtatcagtgaaacccctcatatgcaacccctcgtgtacacgcttctcgtgtgcagctg 1740
Db 1681 AGGTATCAGTGAACCCCTCATATGCAACCCCTCTGTGACACGTTCTCTGTGGCATGACGTG 1740
QY 1741 ctgtgcgtcggggtgagcgttctcagcgtcgcgcgcaagaaatcaagacgttgaggtcgt 1800
```

Db 1741 CTGTCGTCGCGGCGTACGACCTTTTACGCTCCGCGCAGGAATTCAGACCTTGAGGCTCG 1800  
Qy 1801 ccgagacgacgagtagaccgctccggttggtgtgtccactgctgactcaagagctgtt 1860  
Db 1801 CCGGAGCGACGGGTAGACCCGCTCGGTTGGGTGTGCACATGGTGCTCAGCAAGGTCGTTT 1860  
Qy 1861 ttgggtgtcagatgagacctcgaatcgccggttcatagaagctgaagttatgtgacgacg 1920  
Db 1861 TTGGGGGTTCATGATGAGACCTCGAATGGCCCGGTCTATGACCTGATGAGTGGGCGACAC 1920  
Qy 1921 tgcgtctgaagaagaagatttttaacggctcgaagtcgaagaaggcgcgaacacgac 1980  
Db 1921 TGCCTCTTGAGAACAGATTTTACGGCGCTCAGGTGAGAGAGAGGCGCCAGATCGAAC 1980  
Qy 1981 ggcgaagctgagagacacgcggtgtgtctcagcgaagctcacaagggtttagcagcttgaag 2040  
Db 1981 GGCAGGCTGAGGACACACGGGCTGTCTCAGCGACGCTCAGAGGTTGGCCAGCTTGAGG 2040  
Qy 2041 ccgagcgaagcagcagcaaccgacatgcaacccttgaaagcgcggtcccgagcagcggt 2100  
Db 2041 CCGGCGAGCGACGACACACGACACTGGCACCCGTTGAAAGCGCGGCTCCGCGACGGCGGT 2100  
Qy 2101 ggtgacgagctgtgtgtagcgtgtagatgctgtgctgtgtagcagagtggtgtgtagcgtg 2160  
Db 2101 GGTGGACGCTGTGTGAGCGGTGATGAGATCGCTGTGGCTGACAGATGCTGTGTGCGAGTGGC 2160  
Qy 2161 aacagctgtcgaaccagtgtagcagcagaagaagatcgctgtggtgtgtgtctcgtcgtc 2220  
Db 2161 AACATTTGCTGACACAGTGTGAGAGCAAGAGAGATCGCTGTGTGTGTGTGTGTGTGTGTGT 2220  
Qy 2221 agtgcgtcgaagactgtgcggt 2280  
Db 2221 AGTTGTGTGAGAGACTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280  
Qy 2281 agtccaagactgtgtccaagcagcagccctcgtgtgtagaagtgtagtctcgaactgtgta 2340  
Db 2281 AGTTCAAGAGCTGT 2340  
Qy 2341 actgagacgagcagcagaatgtcagctgtgcgaagtgtagtgtgtagcagaagcgcgtcaag 2400  
Db 2341 ACTGAGACCGCGCGGAGAGATGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400  
Qy 2401 agtgcgtcgttagcgcgcgcgaagaatgtgtagtgcgtgtgtgtgtgtgtgtgtgtgtgtgt 2460  
Db 2401 AGT 2460  
Qy 2461 aggtacattccggcgctgt 2520  
Db 2461 AGGTACATTTCCGGGCGCTGT 2520  
Qy 2521 gctcgtgtagctgtgtacgt 2580  
Db 2521 GCGT 2580  
Qy 2581 ccgagcttcattgtccagcgcgcgaagcgtgaagaagcagaagccgagtttagcgtgtgtgtgt 2640  
Db 2581 CCGAGGTTCAATGCCAGCGCAGCAGCTGAAGAGACCAAGCGAGATTAGCCGTCTAACCC 2640  
Qy 2641 gtagagctgtgtccgt 2700  
Db 2641 GTATAGCTTTGCTCCGT 2700  
Qy 2701 ggcgcgaagcgcctacgcggt 2760  
Db 2701 GCGGCGAAGCGGCTACCGGTTTGGGACAGCCCAATTAATCAACCGCTTCGGGTGTGTGAAG 2760  
Qy 2761 cgatcgtgtgtcaagcgt 2820  
Db 2761 CGATGCTGT 2820  
Qy 2821 aagttttttgtatagcgttaagcagcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880  
Db 2821 AAGTTTTCGTATACGCTAATGCCATGATGAGCATCTACTGACAGCGGCAAGCCCGCTCAC 2880

Qy 2881 caacgagcagatctcaggtcgt 2940  
Db 2881 CAACGAGCAGATTTACAGGATGTGGGACAGAGGCGGAGGCTGACCTCCCAACT 2940  
Qy 2941 ccccaagcagc 3000  
Db 2941 CCCCAAGCCACCGCGCGGAGCGCCGCCCTAGAGAGCGGTCCGGGACCGGTACCCGT 3000  
Qy 3001 ggtgtcgc 3060  
Db 3001 GGTCTCGACGGGCGCCACCGTTGGCTCTCAGAGACGACGAACGCGGAGGCTTCAC 3060  
Qy 3061 gaacggttcagacgc 3120  
Db 3061 GAACGTTTACAGCGGATTCGAGCGGACATCCAGATGTGACAGGCTGTGACCTCC 3120  
Qy 3121 acgactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3180  
Db 3121 ACGACTCAGCAGCAGCAAGCACTACCAAGAGACCGGCTCGACGACAGCGCGTGTCTACG 3180  
Qy 3181 cggcaccacagcttcacatccgc 3240  
Db 3181 CGGCAACCCACTTTCTCAACTCCGGCCACTCGACGACGAACAGACCGCGCGCTGGC 3240  
Qy 3241 tcatgactggaacgc 3300  
Db 3241 TCATGATCGGAACCGACCGACGAGCGGCTACTCGAATCTGTGTGTGTGTGTGTGTGTGTGTGT 3300  
Qy 3301 acggtcagcagctgtatcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3360  
Db 3301 ACGGCTACGAACTGATCATCCACGCAATGAAGCCCGCACCAATACCTGACACAGCTCT 3360  
Qy 3361 aaccaagaagaagactgt 3420  
Db 3361 AACCAAGAAAGAACCTGT 3420  
Qy 3421 atcttcagcgt 3480  
Db 3421 ATCTTGACAGT 3480  
Qy 3481 gatgaagcgcgt 3540  
Db 3481 GATGAAGCCGCTGT 3540  
Qy 3541 caagagtggtgc 3555  
Db 3541 CACGAGGTGCGGC 3555

RESULT 3  
AF291751  
LOCUS AF291751 3555 bp DNA circular BCT 01-FEB-2001  
DEFINITION Propionibacterium freudenreichii plasmid p545, complete sequence.  
ACCESSION AF291751  
VERSION AF291751.1 GI:10953186  
KEYWORDS  
SOURCE  
ORGANISM  
Propionibacterium freudenreichii.  
Propionibacterium freudenreichii.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;  
Propionibacterium.  
REFERENCE  
1 (bases 1 to 3555)  
Jore,J.P., van Luijk,N., Luiten,R.G., van der Werf,M.J. and  
Pouwels,P.H.  
Efficient transformation system for Propionibacterium  
freudenreichii based on a novel vector  
JOURNAL Appl. Environ. Microbiol. 67 (2), 498-503 (2001)  
MEDLINE  
2 (bases 1 to 3555)  
Jore,J.P.M. and van Luijk,N.  
Direct Submission  
JOURNAL Submitted (31-JUL-2000) AMGT, TNO Nutrition and Food Research,





[illegible]

QY	1373	g c g a g g c g g a g g t t c c g t g g a c c g t g c a c t a c g c g t a a c a a g a t c g a a c t g a c	1432
Db	2402	A A C G A G G C G G G T G T G C G T T G C G A C C G T T C A C A G C G C T C T A A A G A C C A G C C G A G S A C	2461
QY	1433	g c a t g a	1438
Db	2462	G A A T T C A	2467

RESULT	5
AB007909	
LOCUS	AB007909 6668 bp DNA circular BCT 15-NOV-2000
DEFINITION	Protonibacterium acidipropionici (strain:E214) plasmid pRC01, complete sequence.
ACCESSION	AB007909
VERSION	AB007909.1 GI:11182356
KEYWORDS	Resolvase; Replication protein.
SOURCE	Protonibacterium acidipropionici (strain:E214) plasmid:pRC01 DNA
ORGANISM	Protonibacterium acidipropionici

[illegible]

**TITLE** Yamashita, M. and Murooka, Y. Characterization of pPGOL, a plasmid from propionibacterium acidipropionici, and its use for development of a host-vector system in propionibacteria  
**JOURNAL** Appl. Environ. Microbiol. 66 (11), 4688-4695 (2000)  
**NUMBER** 05060786

JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL	Appl. Environ. Microbiol. 66 (11), 4688-4695 (2000) 20509786 2 (bases 1 to 6868) Murooka, Y., Hashimoto, Y., Kiatpapan, P. and Yamashita, M. Direct Submission Submitted (13-OCT-1997) Yoshikatsu Murooka, Graduate school of
--	--

**JOURNAL**  
Submitted (13-OCT-1997) Yoshiaki Murooka, Graduate school of Engineering, Osaka University, Department of Biochemistry, 2-1 Yamadooka, Suita, Osaka 562, Japan  
(E-mail:murooka@res.bio.eng.osaka-u.ac.jp, Tel:81-6-879-7416, Fax:81-6-879-7418)

FEATURES	Location/Qualifiers
source	1. .6868

```

/organism="Propionibacterium acidipropionica"
/plasmid="pRCO1"
/strain="E214"
/db_xref="taxon:1748"
complement(18..511)

```

```

/note="Orf4"
/codon_start=1
/transl_table=1
/protein_id="BABI7916.1"
/db_xref="GI:11182357"

```

/translation="MDSDPVWALDEEDASALRACHCHGAPVEAHGGRRGPRWCSDDCI  
 KRSEERRAAQAEARPVRTVTVTRHLEGGEEERLEARTVLEELERQALHRODKVKA  
 RHRYGSLMYNDQVFAARRISPADPEGLDLPDQPIEPAASTPTPPPPSEAPFTT  
 VFEHDVSGLGQVPAPQLSGNRAQRAAAHARRH"  
 1075. . 1308  
 CDS

```

/note="orf5"
/codon_start=1
/transl_table=1
/protein_id="BABI7917.1"
/db_xref="GI:11182358"

```

```

/translation="MTATATGKKSSEFGKRIKDELATATGSKSYLREALEDHIDQM
HDVAIRLADVDVGRAGTAYSADVEDDILGDD"
complement(1620..2258)
/note="orf6"
/codon_start=1

```

```

/codon_start=1
/transl_table=1
/protein_id="BAB17918.1"
/db_xref="GI:11182359"
/translation="MESNGSRLAAVARYLSCGRRVTYVRSLLKAENGCSYDAAMAFLLRL

```

gene  
/translation="MESGSLAAVAARYLSCAGRRVRSIKAAQSGSTDAAAEFLIR  
VSAVSOTPPAPDLSPEPTGLVMAAMRQASAPREFCARQADGQAAVAEALAD  
QHRAEFAARDDQAEARAEYOQOEALGLGYQAAVEALTELGAAREDAAREDAQAO  
AEARAEARERQADAAATADTLREITVGLGCGKDTPPAKTTTKTAGRAARS"  
2696..3604



```

CDS
    /gene="repa"
    /gene="repa"
    /codon_start=1
    /transl_table=11
    /product="Replication protein"
    /protein_id="BAB17919.1"
    /db_xref="GI:11182360"
    /translation="MATVTVDEQEWVLPMPPLASDILDAAGIYRMARPALGVRYIE
    VNPQALSNLVVDCDHPDAMRAVMDRHMLPNAIYENPDNGHVAHMLAEALIPTE
    VAHRKPIAAVAATVTEGLRRSYGSDASAGLITRKNPMPANTWCTDHLRLAELTH
    LDAAGLMPAPSRMRRTRRRPPGIGRNCAIFETARTAYADARIIRRHETPRADESD
    LHAVALSTVEALNAGTSEPLPAREAGIASIHRWITHTFTGWDISHVTNEATFTTIO
    STRGRGAGKAPRRARRASTDWEA"
    3604..3948
    /gene="repB"
    /gene="repB"
    /codon_start=1
    /transl_table=11
    /product="Replication protein"
    /protein_id="BAB17920.1"
    /db_xref="GI:11182361"
    /translation="MADVOHRYKRRGTARAEARVCASTRQARWTSIREETITOKA
    VEREIRAKIKDGHGTWGETSRHFGIAKTTAQRARRARRERAEKAEAEALR
    PTLFEGQEGGSA"
    6132..6677
    /codon_start=1
    /transl_table=11
    /product="Resolvasase"
    /protein_id="BAB17921.1"
    /db_xref="GI:11182362"
    /translation="MSITGVARYSTAEOHAAQIDALKAAGADPIYIDKASGATMARP
    OVRACIAGAGPGDILMTVTRIDRLGRSLADLVATDLAGRGVGFSLAEOLDITSPG
    LALFQAGVAFEEEROLITARRPDGLAARARHRRGRPALTAEOQASARQMRAGA
    SLGSTRILIGVSTRRTTRTYT"

BASE COUNT      1321 a      2495 c      1993 g      1059 t
ORIGIN

Query Match      3.3%; Score 118.2; DB 1; Length 6688;
Best Local Similarity 53.2%; Pred. No. 8.1e-11;
Matches 327; Conservative 0; Mismatches 273; Indels 15; Gaps 3;

QY 280 cgttcgagagcgtgtccctcgagagcgtgtgtccgacgacgagcgtgtgtccgacgagc 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2709 CGGTGATGAGCAGTGGGACAGAGTGCTGCTGCCCGCTGCGCTGCTCGACACAC 2768

QY 340 agtcctgggacgtacgagcgtgtgtgtgagagggcgtgtgtgtgtgtgtgtgtgtgtgt 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2769 TGGCAGCGGGCATCTACCGGATGGCCCGCTCGCGGTGGGGGCTCGATACATCGAG 2828

QY 400 cgaaccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2829 TCAACCCCAAGCATAGCAACCTCTCGTGTGCTGCACTCGACACCCGACCGCTGCCA 2888

QY 460 gggcgcgagacgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2889 TGGCGCGCGTGTGGAGACCGCAGACTGGCTGCCCAACGCCATGTGAGAGAACCCGACACA 2948

QY 517 cgaaccgagacacatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2949 ACGGCCACGCGCGCTGTGGGCTGTGGAGACGACCATCCCGACGACGAGTACGCC 3008

QY 577 gggcgcgagacgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3009 ACCGCAAGCCCATCTACCGCGCGCTCGACCGAGGGCGCTCGCATCCCTGACG 3068

QY 637 ggcgtatcctatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 696
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3069 GAAAGCGCTCTCTACGCGCGGCTGATACCAAGAACCCCGACACCCGCTGGAGACCA 3128

QY 697 tctgtggcccgagacgc-----gctctcagagctgtgtgtgtgtgtgtgtgtgtgtgtgt 750

```

```

Db 3129 CCTGTGTCACCGACACACCTCTACCGCGTGGCGGAGCTCGACACCCACCTGTGATCCCGC 3188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 751 acgagatccacgacgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 810
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3189 GCCATATGCCCCGCCCTCTCTGCGGACGACCCCGCGGACCCCGTGGCC----- 3242

QY 811 tcggcgcgacgtcaccctgtctgcacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 870
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3243 TGGCGCCCACTGCGCATCTTGAGACCGCGCGCACCTGGCGCTTACCGGACGCCGCC 3302

QY 871 actctgtggcgcc 885
    |||||
Db 3303 GCATCCGACCAACGCC 3317

RESULT 6
PUE250233
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

PUE250233
6868 bp DNA circular BCT 05-MAY-2001
Propionibacterium jensenii plasmid pLME106 ORF1, ORF2, ORF3, ORF5,
ORF6 and ORF7, ORF8, ORF9 and pna.
AJ250233
AJ250233.1 GI:7635898
invertase; ORF1; ORF10; ORF2; ORF3; ORF5; ORF6; ORF7; ORF8; ORF9;
Pna gene; theta replicase.
Propionibacterium jensenii.
Propionibacterium jensenii.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;
Propionibacterium.
1 (bases 1 to 6868)
Miescher,S., Stierli,M.P., Teuber,M. and Melle,L.
Propionidin SMI, a bacteriocin from Propionibacterium jensenii DFL:
isolation and characterization of the protein and its gene
Syst. Appl. Microbiol. 23 (2), 174-184 (2000)
20383893
2 (bases 1 to 6868)
Stierli,M.P., Melle,L. and Teuber,M.
Molecular analysis of plasmid pLME106, a plasmid isolated from
Propionibacterium jensenii
unpublished
3 (bases 1 to 6868)
Stierli,M.P.
Direct Submission
Submitted (12-OCT-1999) Stierli M.P., Laboratory of Food
Microbiology, LFO G16, ETH Zurich, Institute of Food Science,
Schmelzbergstrasse 9, CH-8092 Zurich, SWITZERLAND
Location/Qualifiers
1..6868
/organism="Propionibacterium jensenii"
/plasmid="pLME106"
/strain="DF1"
/isolate="Raw Milk"
/db_xref="taxon:1749"
/country="Switzerland"
155..388
/note="ORF8"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB88394.1"
/db_xref="GI:7635899"
/db_xref="SPTREML:091459"
/translation="MTLAVSVKSEEGRLDELARATGSKSYLREALEHDIDQV
HDAIARLDADVAGRAATYSADEVDDIIGLDD"
372..635
/note="ORF10"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC38044.1"
/db_xref="GI:1392561"
/translation="MAWTEYTPAVKALRLKLDRAQARRITAYIRLEITGLIDPHQRK

```







Db 4871 GACAAAGCGGCTCTATTGGTGGCCCTAAGGGTGGCATGCCACGCGAATCGTCGCCCGC 4930  
Qy 1147 cccata-----cccgtaacagttcttgagggtctctcatgac 1185  
Db 4931 GTGACAAATTGTCAGAGAGTGCAGGCTGATTCAGAGAGGAGGACTCGATGACATGAC 4990  
Qy 1186 caacagtgaaagctctcccccgaacggtctacgcatcgccgtctgctgcgaaaaagctcgg 1245  
Db 4991 GCGTGAACATGACACCTCGCAATGGATGAGTGGCGCGAGCTGGCGCGCAAGGTCGG 5050  
Qy 1246 tctctcgaagtcacacgctcaagcgggttgacttcggccacacggaggaggtcgtgccc 1305  
Db 5051 GGTGAGCAGAGAAAGACATCGCTCGTTGACCTCGAGCGCGCTGAGGTTTATCTAAGCCG 5110  
Qy 1306 cgttcgcagcagccagcgcgagatctgtagctccgctcgcagagggtcag-agcatgctg 1364  
Db 5111 TGCCGCAACAGCGCGCTGAGCGGATTCGTGAGCTGCGCAAGGCGCTGCAATGCGTG 5170  
Qy 1365 cgatgctgcgcagaggttccgttggcagcagctgcaactagcgcgtgaagaagatc 1424  
Db 5171 CGATGCGCGGTGAGGCTGAGTGTCTCGTTGGCACCCTCCACAAACGCGTCAAAAGACACG 5230  
Qy 1425 gaactgacg 1433  
Db 5231 CCGAGGACG 5239

RESULT 8  
AF312210 5403 bp DNA circular BCT 13-DEC-2000  
LOCUS Rhodococcus erythropolis plasmid pN30 replication protein and  
DEFINITION putative DNA-binding protein genes, complete cds.  
ACCESSION AF312210  
VERSION AF312210.1 GI:11095364  
KEYWORDS  
SOURCE Rhodococcus erythropolis.  
ORGANISM Rhodococcus erythropolis.  
Bacteria; Filumetes; Actinobacteri; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
REFERENCE  
AUTHORS Ryabchenko, L.E., Novikov, A.D., Golyslin, P.N. and Yansenko, A.S.  
TITLE Rhodococcus erythropolis plasmid pN30 putative replicase (ORF1),  
and putative DNA-binding replication protein (ORF2)  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 5403)  
Ryabchenko, L.E., Novikov, A.D., Golyslin, P.N. and Yansenko, A.S.  
TITLE Direct Submission  
JOURNAL Submitted (11-OCT-2000) Biodegradation Genetics Laboratory,  
Institute of Genetics and Selection of Industrial Microorganisms,  
1st Dorozhny Pr., 1, Moscow 113545, Russia  
FEATURES  
source  
1. 5403  
/organism="Rhodococcus erythropolis"  
/plasmid="pN30"  
/db\_xref="taxon:1833"  
1160..2047  
/note="ORF1: putative; similar to putative Rhodococcus  
rhodococcus pRC4 putative replication protein (repA),  
Rhodococcus rhodochrous pKA22 putative theta replicase  
(ORF2). Rhodococcus erythropolis pFAJ2600 putative  
replication protein (repA), Mycobacterium fortuitum  
pAJ5000 putative replication protein (37K)."  
/codon\_start=1  
/transl\_table=1  
/product="replication protein"  
/protein\_id="AA029854.1"  
/db\_xref="GI:11095365"  
/translation="MKSATGVLSDROMADAVLLPHRFATNNLQROYSRSDALAL  
MRVESHPLALIGSTVIDCDHVDAMRAFEPSHPAPNNVAQSPSGRAHGMWIGPN  
HVCRTDARLPLRYARIETGLKISVGDAVAGOLKNDIHPDWETIGPAPYTL  
ROLATITRPQMPRPDRPRAVIGLGNVMPATPRWAPQMOHONGCRMDHVLQH  
CHAVTEFTTPLEPTVEYRATQSIKIMWRTTEQYRARAHLGKGGKATTLAKOE  
AVRNNAKKYDEHTMREAIT"

CDS  
2047..2331  
/note="ORF2: putative; similar to Rhodococcus erythropolis  
pFAJ2600 putative DNA-binding replication protein (repB),  
Rhodococcus rhodochrous pRC4 putative DNA-binding protein  
(repB)."  
/codon\_start=1  
/transl\_table=1  
/product="putative DNA-binding protein"  
/protein\_id="AA029855.1"  
/db\_xref="GI:11095366"  
/translation="MGGAKNPRKMTAAAAAEFGASTRTQRLFAEPBDYLRK  
ARRKAYELRKQGLKRIEIAEMLELSTGRLHLDARHGEISAEIDLTA"

BASE COUNT 1073 a 1719 c 1647 g 964 t  
ORIGIN

Query Match 2.7%; Score 95.8; DB 1: Length 5403;  
Best Local Similarity 49.2%; Freq. No. 6.1e-07;  
Matches 435; Conservative 0; Mismatches 402; Indels 48; Gaps 5;

Qy 549 aaacctgtgtctcagatcgatcgatcgagcgcgagcctatcaactgtctgcgcgcgtc 608  
Db 1463 AACCAAGTGTGCGCGACGACAGCGCCCGGACTGAGACCTACGCTAGCCGACCGCATC 1522  
Qy 609 gaggcagcctatcgagcgtctcgcgcgcgagatcctcctacgagccagcagcaag 668  
Db 1523 GAACCGGCTCAGATCAGCGCTCGCGGCGGATTCGGGTATGCGGCGCACTGACCAAA 1582  
Qy 669 aaccgctcagacagccagcctatcgagcgcctctgagggcccgagagcgtctagaagctg 728  
Db 1583 AACCCGATTACCCCGATTGGAGACGATCTAGGCGCCGCGCACCCCGCTACACATTTGGG 1642  
Qy 729 cgcgcctcgcacacacctcgcagagatccagcagctcgcgcgcgcgcgcgcgcgcgc 788  
Db 1643 CAG-----CTGCCACATCCACACACCCCGCA-----GATGCCGCT 1681  
Qy 789 cgcaagtcacccagatcaagcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 848  
Db 1682 CGGCGCGCTCGGCGCGCTGCGGCGCGCAACGTCACCATCTGAGCGCACCGCGCGGA 1741  
Qy 849 tggagatccgc 908  
Db 1742 TGGCATACCCGCAAGTGTGCGACACCGAAGAACGCGCGCGGCGGACTGGGACCATCTC 1801  
Qy 909 gttatcgacatccactaetgaagcagagatcatcgcgcgcgcgcgcgcgcgcgcgcgcgc 968  
Db 1802 GTTCGACGACATGCGCACCGCTCAACACCGAGT-----CACGACA 1843  
Qy 969 cccctcgcgttgaacgaacttaagcactatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1028  
Db 1844 CCACGCGCTTCACCGAGTACGCGCCACCGCGCAATCCATCTCCAAATGATGTGGCGC 1903  
Qy 1029 aactcaccgccgaactctcgc 1088  
Db 1904 AATTTACCGAAGACAGTACCGAGCCGAC--AAAGGCAATCTGTGTCAAANAAGCGCGC 1960  
Qy 1089 aaagcgcgcaagaagc 1148  
Db 1961 AAGGCAAGGACACTGCGCAACCAAGACCGCTCCGCAAAACATGCAAGAAAGTACGACGA 2020  
Qy 1149 cataccgtcaacagttcttgagggtctcctatgacacacgltgaacgtctcccgca 1208  
Db 2021 CATACGATGCGAGA-----GGCGATWTGTCGATGGCGGAGCCAAATAATCCGGTGGCC 2074  
Qy 1209 acggttaagcatcgcgctctcgc 1268  
Db 2075 GAAGATGACGCGACACACACACCGCAAAATTCGCTCCACTGCAACATTCACAC 2134  
Qy 1269 gttgacttcgagcagc 1328  
Db 2135 GCTTGTTCGTCAGCGCGGCTGACGATTCCTCGCGCGGCAAACTGCGGTGCAAG 2194  
Qy 1329 ttcgtgaactcgcgtcgc 1388

Db 2195 CTGTCAGCTGGGAGGAGGAGGAGTACCGGGAATTCGCGAAGAGCATGAACTCT 2254

Qy 1389 ccgtggagccgctgacacgacgctgacacgaagatgaacgacg 1433

Db 2255 CGACGGGATGCTGCGGCGATTTACTGACGAGCGCGGAGGCGACG 2299

RESULT 9

AB040101 2582 bp DNA linear BCT 06-APR-2001

LOCUS Rhodococcus rhodochrous plasmid pRC4 repB genes for

DEFINITION replication protein, DNA binding replication protein, complete cds.

AB040101

AB040101.1 GI:7262572

VERSION DNA-binding replication protein: replication protein.

KEYWORDS Rhodococcus rhodochrous (strain:IFO3338) plasmid:pRC4 DNA.

SOURCE Rhodococcus rhodochrous

ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.

REFERENCE 1 (sites)

AUTHORS Hirasawa,K., Ishii,Y., Kobayashi,M., Kozizumi,K. and Marubashi,K.

MEDLINE Improvement of desulfurization activity in Rhodococcus erythropolis

AUTHORS Blosci, Biotechnol. Biochem. 65 (2), 239-246 (2001)

TITLE 21196497

JOURNAL 2 (bases 1 to 2582)

21196497

Submitted (14-MAR-2000) Kenji Marubashi, Petroleum Energy Center,

Direct Submission Bio-Retining Process Laboratory; Sodeshi-cho 1900, Shimizu,

Ishii,Y., Hirasawa,K. and Marubashi,K. Shizuoka 424-0037, Japan (E-mail:k.marub@prpl.pecj.or.jp,

Tel:+81-543-67-9550, Fax:+81-543-67-9552)

FEATURES

source location/Qualifiers

1. 2582

/organism="Rhodococcus rhodochrous"

/plasmid="pRC4"

/strain="IFO3338"

/db\_xref="taxon:1829"

1142. 2062

/gene="repA"

1142. 2062

/note="repA"

/note="putative

similar to putative theta replicase from Rhodococcus

rhodochrous PKA22 and RepA proteins from Mycobacterium

fortuitum pAL5000 and related plasmids."

/codon\_start=1

/transl\_table=11

/product="replication protein"

/protein\_id="BAA92818.1"

/db\_xref="GI:7262573"

/translation="MDMSGGSGSDGMDLMLPLPLANDLLGVYRMPDADRY

LEANPQALSNLVYDHPDALRLSLAAGNHPLENAIYENRNGHAHVAALTEPT

REIYARKPLAAVAVNGELRAVVDGDAVSLMTKNPHSDMTHMHAERSLADL

EHDLKHPPEPRMQSKRRREDVGLGRMCMLFETARTAYBELCHMDPDGLKAI

OYEADILNAAFEPLPVSVEIRAIAASHIRWITVTKSRMADGPAVEATFVAIOSARG

KWTEKKREANRRATKYDRDLVRKEATGSS"

2052. 2333

/gene="repB"

2052. 2333

/note="repB"

/note="putative

similar to putative DNA-binding replication protein from

Rhodococcus erythropolis pRM2600"

/codon\_start=1

/transl\_table=11

/product="DNA-binding replication protein"

/protein\_id="BAA92819.1"

/db\_xref="GI:7262574"

/translation="MGAEPPARTRTAPRVARVARGASPTVRRITAEPPAARAA

RRKOYLELRASMKRREIAAEVAGSVGGITLHARKTEQSKAGANA"

gene

CDS

BASE COUNT 524 a 828 c 755 g 475 t

ORIGIN

Query Match 2.7%; Score 95.4; DB 1; Length 2582;

Best Local Similarity 48.6%; Pred. No. 8.1e-07;

Matches 295; Conservative 0; Mismatches 306; Indels 6; Gaps 1;

Qy 301 agagctgctgctgacacgaacgctgctgacgacgaagaagcttgaggctgacgacgacg 360

Db 1182 AGTTGTGGCTGCTCTGTGTGGCCGCTCGCAACGGACGATTTGTTGGTGGGCTTACCGGA 1241

Qy 361 tgactgagcagagggcgctgagagctgcttaacatgaagcgaacccgttgtaicagc 420

Db 1242 TGCCTCGCAGGATGGGCTCGATCGGCGCTACCTTGAGGCCAATCCGACGCGCTAGCA 1301

Qy 421 ccctgtgctacacgctgacgacgctgacgacgctgacgacgacgacgacgacgacgacg 474

Db 1302 ATCTCTCTGCTGCTGATGTCGATCATCAACGCGGACATGCGGCTCTGTCTGCGCGC 1361

Qy 475 ctgagctgcttaccgcttccacgctgcttccacgctgcttccacgctgcttccacgctg 534

Db 1362 GCACCATCCCTTGCCGGAACGGATGCGAAGAACCCGCGCATGGACACGACATGCGG 1421

Qy 535 tctatgcttgaagaacccctgtgtctgacgacgacgacgacgacgacgacgacgacgac 594

Db 1422 TGTGGGCAATGACCGAACCCTTTCACGCGCACGAGTACGCAAGCAAGCAAGCAAGC 1481

Qy 595 tgctgacgacgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 654

Db 1482 ATGCCCGACGGGTAAAGAGAGAGGCTGCTGACCTGCTGATGCGGATGCGGATGCGG 1541

Qy 655 accgagacacgaagaacccgctgacgacgacgacgacgacgacgacgacgacgacgacg 714

Db 1542 GGTGATGAGAGAGAACCCGACACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 1601

Qy 715 cgcttcaagagctgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 774

Db 1602 CTGATCGCTGGCGGATCTTCGAAATGATGATGATGATGATGATGATGATGATGATG 1661

Qy 775 caaggaacccgctgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 834

Db 1662 GACAGAGCAAACTGCTGCGGAGAGCCAGTCCGACCTGACCTGACCTGACCTGACCTG 1721

Qy 835 acacaccccgatgtggaacgacgacgacgacgacgacgacgacgacgacgacgacgacg 894

Db 1722 AGACGGACGCACTTGGCGCATACCGGAATTGGGTTGCCATTGGGAGATCCGGAAGTT 1781

Qy 895 aatggga 901

Db 1782 TAGGGAA 1788

RESULT 10

BL039878 1746 bp DNA linear BCT 15-JUN-2001

LOCUS Brevibacterium linens plasmid pRB1 putative theta replicase

DEFINITION (orf310) gene, complete cds.

AB039878

AB039878.1 GI:1079741

VERSION U39878.1

KEYWORDS Brevibacterium linens.

SOURCE Brevibacterium linens

ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococccineae; Brevibacteriaceae; Brevibacterium.

REFERENCE 1 (bases 1 to 1746)

AUTHORS Ankl,S., Bouvier,I., Reyes,O., Predali,F. and Leblon,G.

TITLE A Brevibacterium linens pRB1 replicon functional in

JOURNAL Corynebacterium glutamicum

MEDLINE Plasmid 36 (1), 36-41 (1996)

PUBMED 97092385

8938050

REFERENCE 2 (bases 1 to 1746)

AUTHORS Ankl,S., Bouvier,I., Reyes,O., Predali,F. and Leblon,G.



[illegible]

REFERENCE	LOCUS	ACCESSION	VERSION	KEYWORDS	ORGANISM
1 (bases 1 to 6407)	Sequence 2 from patent US 5583038.	131848	131848	GI:1822639	Unknown.
Bacterial expression vectors containing DNA encoding secretion signals of lipoproteins					Unclassified.
Patent: US 5583038-A 1 10-DEC-1996;					
Location/Qualifiers					
1..6407					
/organism="unknown"					
BASE COUNT	1399 a	1800 c	1770 g	1438 t	
ORIGIN					
Query Match	2.3%	Score 82.2;	DB 6;	Length 6407;	
Best Local Similarity	47.8%;	Pred. No. 0.00013;			
Matches 274;	Conservative 0;	Mismatches 293;	Indels 6;	Gaps 1;	
301	agaagctgctccacgaagccgctgtgcgtcaagccgaagatcttgggagctacggcgaag	360			
2852	AGCTGTGGCTGCCGACCTACGCGCGGTGCGCAAGGAGATCTCTCGAGGGGATCTACCGCC	2793			
361	tgactcgcgacagaggcgctgtgagctgctctcatatgaagcgaaccgctgtgtcatgcagt	420			
2792	AAAGCCGGCGGCTCGGCGCTTAGAGCGCCCGGTACATGAGGCGCAACCAACAGCGCTGC	2733			
421	cccttgatcatcacccgatcagatgcttgcgtgagtgcgtgacgtggccgagacctgctgggc	480			
2732	ACCTGCTGTGTGAGACGTAAACATCCAGACGACGCGCTCCGAGCGCTCAGCGCCGG	2673			
481	tgcttcacgctcctcagctgttcacgt-----aacgctgtcagcaccacccgagacacatcg	534			
2672	GGTCCATCCGCTCGCCCAACGCGATCTGGGCAATCGGCAACGCGCAACGCGCACGACGAG	2613			
535	tctatgctctgaagaacctgtgtgtctgcacccatgagccgcgcgagagcgtctataacc	594			
2612	TGTGGGCACTCAACGCCCTCTGTCCACGCAACCGAATACGCGCGGTAAAGCGCTCGCAT	2553			
595	tgctgcgcgcgctgcagagccctatgagcgttctcggcgcgatgcatcctacgggc	654			
2552	ACATGGCGGCGGTGCGCGCAAGGCTTCGCGCGCGGTGACGCGCAACCGCATTTACTCAG	2493			
655	accggtatcaagaagaccgcgctcaagcaccgccatgagcctctggggcccgagacg	714			
2492	GCTCATGACCAAAAACCCCGGCGACATCTGCTGGGAAAGGAATGGCTCCACTCAGATC	2433			
715	cgctctcagagctgcgagccctcctgcacacacccctgcagagatccaagcaactgcggaag	774			
2432	TCTAACACTCAGCCACATCTGAGGCGGAGCTGCGGCGCAACATGCGACCGCGCTCTGGC	2373			
775	cagggaaccgcgctgcgaagctaccgcgataaagtgctggcggaagctaacccgttgc	834			
2372	GTCAGCAACACACGTACAAAGCGGCTCCGAGCCCGCTTAGGGGGAATTGGCAGCTGTTGC	2313			
835	aaacacaccgcgatgtgagcatcacccgagccgtcc	867			
2312	ATTCCGTCAGGTTGGGGCTATCTCCCGGCC	2280			
RESULT 13					
LOCUS	131848	6407 bp	DNA	Linear	PAT 06-FEB-1997
DEFINITION	Sequence 2 from patent US 5583038.				
ACCESSION	131848				
VERSION	131848.1	GI:1822639			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 6407)				
AUTHORS	Stover,C.K.				
TITLE	Bacterial expression vectors containing DNA encoding secretion signals of lipoproteins				





**THIS PAGE BLANK (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2002, 11:26:07 ; Search time 532.75 Seconds  
(without alignments)  
11456.839 Million cell updates/sec

Title: US-09-720-583A-1

Perfect score: 3555  
Sequence: 1 gtcgaccttaccagccgccc.....ccgcacgagtgccgccc 3555

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3555	100.0	3555	21	AAZ49691
2	82.2	2.3	1382	19	AAV69313
3	82.2	2.3	1463	19	AAV69312
4	82.2	2.3	6407	18	AAV64413
5	80.6	2.3	4120	13	AAQ21502
6	80.6	2.3	4120	18	AAQ21502
7	80.6	2.3	4296	17	AAV64421
8	80.6	2.3	4352	17	AAV64421
9	80.6	2.3	4505	17	AAV64421

10	80.6	2.3	6047	17	AAV64421	Plasmid pMW30-lux
11	80.6	2.3	6171	17	AAV64421	Plasmid pMW261-lux
12	77.4	2.2	4120	17	AAV64421	Plasmid pMW206-seq
13	68.6	1.9	4119	14	AAQ41331	pMW206 - cassette
14	68.6	1.9	6407	14	AAQ41316	pMW101 - a deletion
15	58.6	1.6	4811	13	AAQ31743	pMW101 - a deletion
16	58.6	1.6	4811	13	AAQ31727	pMW101 - a deletion
17	56.2	1.6	4674	21	AAV14666	Nucleotide sequenc
18	56.2	1.6	4725	21	AAV14665	Nucleotide sequenc
19	56.2	1.6	4737	21	AAV14668	Nucleotide sequenc
20	56.2	1.6	4767	21	AAV14667	Nucleotide sequenc
21	56.2	1.6	4818	21	AAV14669	Nucleotide sequenc
22	52.4	1.5	65140	22	AAV17184	Streptomyces nous
23	52.4	1.5	125401	22	AAV17186	Streptomyces nous
24	50.2	1.4	1524	20	AAV09259	Rice RPC213 CDNA
25	50.2	1.4	1524	22	AAV74817	Rice CDNA sequence
26	50.2	1.4	4466	21	AAV14663	Nucleotide sequenc
27	50.2	1.4	4478	21	AAV14661	Nucleotide sequenc
28	50.2	1.4	4547	21	AAV14664	Nucleotide sequenc
29	50.2	1.4	4571	21	AAV14662	Nucleotide sequenc
30	50.2	1.4	77536	21	AAV14651	Nucleotide sequenc
31	49.8	1.4	1881	24	AAV18436	Contig 109 DNA enc
32	49.4	1.4	1350	21	AAV5844	Mitomycin gene clu
33	49.4	1.4	18034	21	AAV5841	Complete Mitomycin
34	48.2	1.4	833	15	AAV64203	snab gene encoding
35	48.2	1.4	1896	21	AAV48297	S. coelicolor YesM
36	48.2	1.4	5392	15	AAV64201	Sequence compisn
37	47.4	1.3	27541	22	AAV17185	Streptomyces nous
38	47.4	1.3	58857	22	AAV58471	Nucleotide sequenc
39	47	1.3	5824	22	AAV44991	CDNA encoding nove
40	47	1.3	5824	22	AAV51846	Human polynucleoti
41	47	1.3	5824	22	AAV51846	Human polynucleoti
42	47	1.3	77536	21	AAV14651	Nucleotide sequenc
43	46.6	1.3	11220	21	AAV87298	S. venezuelae mmcr
44	46.6	1.3	36778	21	AAV87318	S. venezuelae plx
45	46.6	1.3	37948	21	AAV87285	S. venezuelae plx

#### ALIGNMENTS

RESULT 1						
ID	AAZ49691	standard; DNA; 3555 BP.				
XX						
XX	AAZ49691:					
DT	07-APR-2000	(first entry)				
XX						
XX		Proplombacterium plasmid LMG 16545 DNA.				
DE						
XX						
XX		Proplombacterium LMG 16545: vector; plasmid; antigen; vaccine; enzyme;				
KW		nutritional factor; growth factor; clotting factor; antimicrobial; drug;				
KW		hormone; vitamin B12; animal feed; lactic acid bacteria; foodstuff;				
KW		cheese; cyclic; circular; ds.				
XX						
OS		Proplombacterium freudenreichii LMG 16545.				
XX						
FH						
FT	key	Location/Qualifiers				
FT	CDS	273..1184				
FT		/tag- a				
FT		/product- "Proplombacterium LMG 16545 protein-1"				
FT		/note- "Encoded by ORF-1"				
FT		1181..1438				
FT		/tag- b				
FT	CDS	/product- "Proplombacterium LMG 16545 protein-2"				
FT		/note- "Encoded by ORF-2"				
PN	MO9967356-A2.					
XX						
PD	29-DEC-1999.					
XX						
PF	25-JUN-1999.	99MO-EP04416.				

PR 25-JUN-1998; 98EP-0305033.  
XX  
XA (KONN ) GIST-BROCADES BV.  
PA  
XX  
PI Pouwels PH, Van Luijk N, Jore JPM, Lulten RGM;  
XX  
DR WPI: 2000-136977/12.  
DR P-PSDB: AAY44636, AAY44637.  
XX  
XX Novel vectors containing Propionibacterium sequences, used to express  
PT homologous or heterologous proteins -  
PS  
PS Claim 1a; Page 39-43; 54pp; English.  
XX  
XX The present DNA sequence is the endogenous plasmid of Propionibacterium  
CC IM6 16545, derived from Propionibacterium freudenreichii deposited under  
CC CBS 101022 or CBS 101023. This vector has insertion sites for foreign  
CC DNA fragments and is capable of autonomous replication. It can be used  
CC safely as they remain extrachromosomal and are very stable. They can be  
CC used to express homologous or heterologous proteins, which may include  
CC antigens for use in vaccines, nutritional, growth and clotting factors,  
CC antimicrobials, vitamins (especially vitamin B12), enzymes, hormones and  
CC drugs. The Propionibacterium host cells are used in the production of  
CC animal feeds, manufacture of foodstuffs like, sausages and cheese, where  
CC they may be used instead of or in addition to lactic acid bacteria and  
CC in the manufacture of vitamin B12. This sequence can also be used as a  
CC source of primers and probes.  
XX  
XX Sequence 3555 BP; 705 A; 1156 C; 1075 G; 619 T; 0 other;

Query Match	100.0%	Score 3555	DB 21	Length 3555
Best Local Similarity	100.0%	Pred. No. 0		
Matches 3555	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	gtcgacccctgacagccggtcgagcagttcaagcggaagatcgcacagctgcgcgaggaacta	60		
Db 1	gtcgacccctgacagccggtcgagcagttcaagcggaagatcgcacagctgcgcgaggaacta	60		
QY 61	gtcgacatgcccgcgaacacgcgccacagccatcccttggagagcaggtgacagcgctcaaggaggt	120		
Db 61	gtcgacatgcccgcgaacacgcgccacagccatcccttggagagcaggtgacagcgctcaaggaggt	120		
QY 121	cgggggagatgtttggagaggagatgttgagaaaggagatgttcgtctgtgtcccaaatgtctaaacg	180		
Db 121	cgggggagatgtttggagaggagatgttgagaaaggagatgttcgtctgtgtcccaaatgtctaaacg	180		
QY 181	gttaactaactgtatattgggtctctctgtctgcgccacatttaacacacgcgcgaaggaaatgagacac	240		
Db 181	gttaactaactgtatattgggtctctctgtctgcgccacatttaacacacgcgcgaaggaaatgagacac	240		
QY 241	gtctgaacgttgatctcgcatgtcttcaactgtcatgtatagatctgtctcgagacggtttgtccctgt	300		
Db 241	gtctgaacgttgatctcgcatgtcttcaactgtcatgtatagatctgtctcgagacggtttgtccctgt	300		
QY 301	agagatgtgctctgcacgaacgaacgcgcctctgtgctgcgttcgaacgaagaatgtcttggcgatcacgcgcgaag	360		
Db 301	agagatgtgctctgcacgaacgaacgcgcctctgtgctgcgttcgaacgaagaatgtcttggcgatcacgcgcgaag	360		
QY 361	tgactctgcgcagagggcgctgtgagctgtgccttcaatcgaagacgaaacccggtgtgtcatgtacgt	420		
Db 361	tgactctgcgcagagggcgctgtgagctgtgccttcaatcgaagacgaaacccggtgtgtcatgtacgt	420		
QY 421	ccttgggtcatacacgcgatatcgatgtcttcggatgtgtgactgtgactggcgacagacctgcgtgggc	480		
Db 421	ccttgggtcatacacgcgatatcgatgtcttcggatgtgtgactgtgactggcgacagacctgcgtgggc	480		
QY 481	tgacctaacccgtctccacgtgtgtccatggaacccggtgtacagaccccggaacacatcgtcgttat	540		
Db 481	tgacctaacccgtctccacgtgtgtccatggaacccggtgtacagaccccggaacacatcgtcgttat	540		
QY 541	ccttgaagaaacccgtgtgtgtctgacacggtatgcgcgcgcgcgcgaagcgctatacaacctgtctgc	600		

[illegible]





Best Local Similarity 47.8%; Pred. No. 4e-10;  
Matches 274; Conservative 0; Mismatches 293; Indels 6; Gaps 1;

```

QY 301 agagctggtgctgcaagccgctgctgctgcaagagctggtggttaccggcagc 360
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 208 agctctggtgctgctgctgctgctgcaagagcagctgctgctggttaccggc 267
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 tgaactcgagagggcgctggtgagctgcttaccagcaagccgctgtgtacatgca 420
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 aagacgcgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgca 327
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 ccttggtacacacgctgctgctgctgctgctgctgctgctgctgctgctgctgca 480
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 328 accctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 387
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 tgccttcacgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgca 534
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 388 ggtcccatcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 447
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 tctatgcttgaagaacctgtgtgtgctgacagatgctgctgctgctgctgctgca 594
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 448 tgtggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 507
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 595 tgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 654
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 508 acatggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 567
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 accggtacacagaacccgctgctgctgctgctgctgctgctgctgctgctgctgca 714
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 568 gctctatgacacaaaccccgctgctgctgctgctgctgctgctgctgctgctgca 627
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 gcgtctagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 774
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 628 tctacacactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 687
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 775 cagggaacccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 834
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 688 gtcaagacacacgctgctgctgctgctgctgctgctgctgctgctgctgctgca 747
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 acacacccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 867
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 748 attcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgca 780
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
AAT64413/C
ID AAT64413 standard; DNA; 6407 BP.
XX
AC AAT64413;
XX
XX 28-MAY-1997 (first entry)
XX
DE plasmid pmv101.
XX
XX Plasmid pmv101; Mycobacterium; lipoprotein secretion signal sequence;
KW antigen; bacterial vaccine; Lyme disease; Borrelia burgdorferi;
KW surface protein; cyclic; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT complement (1598..2542)
FT CDS
FT /tag= a
FT /label= M.rep ORF1
FT /note= "Supports autonomous replication in BCG"
FT complement (1954..3060)
FT CDS
FT /tag= b
FT /label= M.rep ORF2
FT /note= "Supports autonomous replication in BCG"
FT 5258..6073
FT CDS
FT /tag= c
FT /mod_base= kan
FT /note= "Kanamycin resistance gene"

```

```

FT misc_feature 187..1310
FT /tag= d
FT /note= "Nari/Bali fragment deleted in pMV111"
FT primer_bind
FT /tag= e
FT /bound_moety= Primer Bam-M.rep
FT misc_feature
FT /tag= f
FT /note= "CATG duplication in pMV110"
FT 3221..3860
FT /tag= g
FT /note= "Fragment deleted in M0112-200s"
FT 5144..5164
FT /tag= h
FT /bound_moety= Primer BamMhp-Kan

US583038-A.
10-DEC-1996.
21-OCT-1991; 910S-0780261.
17-NOV-1992; 92US-0977630.
21-OCT-1991; 91US-0780261.
(MED1-) MEDIMUNE INC.
Stover CK;
WPI; 1997-042315/04.
P-PSDB; AAW14834-36.

Mycobacteria expressing secretion signal of lipoprotein and
heterologous antigen, esp. outer surface protein A or B of Borrelia
burgdorferi - are used in the form of a live bacterial vaccines
against Lyme disease

Example 1; Column 27-34; 122bp; English.

This sequence represents the plasmid pmv101. This plasmid was
used in the construction of a series of vectors used to produce
the recombinant Mycobacterium of the invention. The recombinant
Mycobacterium of the invention are transformed with DNA encoding
a polypeptide which comprises a lipoprotein secretion signal
sequence and an antigen (Ag) heterologous to the Mycobacterium.
The lipoprotein secretion signal causes the Ag to be produced as
a lipoprotein. The Mycobacterium may be used in the form of a live
bacterial vaccine against Lyme disease, where the bacteria express
a surface protein of Borrelia burgdorferi, the causative agent of
Lyme disease.

Sequence 6407 BP; 1399 A; 1800 C; 1770 G; 1438 T; 0 other;

Query Match 2.3%; Score 82.2; DB 18; Length 6407;
Best Local Similarity 47.8%; Pred. No. 6.3e-10;
Matches 274; Conservative 0; Mismatches 293; Indels 6; Gaps 1;

QY 301 agagctggtgctgcaagccgctgctgctgcaagagctggtggttaccggcagc 360
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2852 Agctctggtgctgctgctgctgctgctgctgctgctgctgctgctgctgca 2793
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 tgaactcgagagggcgctggtgagctgcttaccagcaagccgctgtgtacatgca 420
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2792 AAAGCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 2733
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 ccttggtacacacgctgctgctgctgctgctgctgctgctgctgctgctgctgca 480
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2732 ACCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGCA 2673
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 tgccttcacgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgca 534
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2672 GGTCCATCCGCTGCCAACGCGATCGTGGCAATGCGCGCAACGCGCACACGACG 2613
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY	535	tctatgcttgaagaacccgtgtgtctgacgatlccgcgcgcgcgaagcctctcaacc	594
Db	2612	TGTGGGCACTCAGCCGCCCTCTTCACGCGCAATACCGCGCGGTAAAGCGCTCGCAT	2553
QY	595	tgtctgcgcgcgcgttcgaagcaggtctatgtcgaagcttctcgcgcgcgatlcatctctacgagc	654
Db	2552	ACATGAGCGCGGTGGCGGAAGCGGCTTCGGGGCGCGCTGCAGCGCAGCGCATGTTACTAG	2493
QY	655	accggatccaagaagaaccgcgtctagacacgcgcgcacatgtgcgaacctcttggggccccgcagacg	714
Db	2492	GCCCTCATGACCCAAAACCCCGCGGCACATGCGCTGGGGAACGGAAATGGCTCCACTCAGATC	2433
QY	715	cgcctcagcagcgtgcgcgccttcgacacacacctcgaagagatccaagcactgtccggagc	774
Db	2432	TCTTACACACTCAGCCACATGTGAAGCCGAGCTCGGCGCCGACATGCGACCGCGCTGGC	2373
QY	775	caagggaaccgcgtcgcgaagtlcaccgcgatlcaacggtcgcgcgcgaacgltcacacctgttcg	834
Db	2372	GTGAGCAGACGACGTACAAAGGCGGCTCCGACGCGCGGTAGGGGGAAATTGGCGACTGTTCG	2133
QY	835	acaccaccgcgaatgtgtgcatccgcggccgtcc	867
Db	2312	ATTCCGTGAGGTTGTGGGCTATCTGTCGCGCC	2280
RESULT	5		
AAQ21502			
ID	AAQ21502	standard; DNA; 4120 bp.	
XX			
XX	AAQ21502:		
DT	03-JUN-1992	(first entry)	
XX			
DE	Vector pMW206 for cloning genes in E.coli and mycobacteria.		
XX			
KW	Polymerase chain reaction: mycobacterial promoter; kanamycin;		
KW	kanR; aph gene; BCG; Bacille Calmette-Guerin;		
KW	site-specific integration; ss.		
XX			
OS	Escherichia coli.		
OS	Mycobacterium smegmatis.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	121..936	
FT		/*tag- a	
FT		/product- kanR	
FT		/note- "kanamycin resistance gene"	
FT		1..1219	
FT		/*tag- b	
FT		/note- "kanamycin resistance cassette"	
FT		1220..2011	
FT		/*tag- c	
FT		/note- "E.rep cassette"	
FT		2012..2025	
FT		/*tag- d	
FT		/note- "MluI-NotI linker"	
FT		2026..3924	
FT		/*tag- e	
FT		/note- "W.rep cassette"	
FT		3925..4016	
FT		/*tag- f	
FT		/function- Multiple cloning site	
FT		/note- "contains 16 unique restriction sites, stop codons in all 3 reading frames and a transcription terminator"	
XX			
XX	WO9201783-A.		
PN			
XX	06-FEB-1992.		
PD			
XX			
XX	09-JUL-1991.	91WO-US04833.	
XX			

PR	16-JUL-1990;	90US-0553907.
XX	(YESH ) EINSTEIN A COLLEGE.	
PA	(UPT-) UNIV OF PITTSBURGH.	
XX		
PI	Jacobs WR, Hatfull G;	
DR	WPI; 1992-064943/08.	
XX	P-PsDB; AAR20993.	
PT	DNA site-specific integration into Mycobacterium - useful as	
PT	adjuvant in vaccines and as therapeutic agent for malaria,	
PT	influenza, herpes and human immunodeficiency virus	
XX		
PS	Example 3; Fig 31; 82pp; English.	
XX		
CC	Cassettes of all the components necessary for plasmid replication	
CC	in E.coli and mycobacteria (E.rep and M.rep, respectively) and for	
CC	selection of transformants (kanr) were constructed using the PCR	
CC	technique. A multiple cloning site was synthesised. The cassettes	
CC	were constructed to allow directional cloning and assembly into a	
CC	plasmid where all transcription is unidirectional. Plasmid pmv206	
CC	was derived from plasmid pmv204 which was made up of all the	
CC	cassettes. It was modified by the insertion of an MluI-NotI linker	
CC	between the M.rep and E.rep cassettes to facilitate removal of	
CC	M.rep in further constructions.	
CC	See AA023568-021573 for the PCR primers used in the construction of	
CC	the individual cassettes.	
XX		
Q0	Sequence 4120 BP; 923 A; 1127 C; 1163 G; 907 T; 0 other;	

[illegible]

Db	2922	atccgcacaggtctgtggcgccatcgtccgcc	2954
RESULT	6		
ID	AAT64421		
XX	AAT64421	standard; DNA; 4120 BP.	
AC	AAT64421;		
XX			
DT	28-MAY-1997	(first entry)	
XX			
DE	Plasmid pMW206.		
XX			
KM	Plasmid pMW101; Mycobacteria; lipoprotein secretion signal sequence;		
KW	antigen; bacterial vaccine; Lyme disease; Borrelia burgdorferi;		
XX	surface protein; primer; amplify; PCR; polymerase chain reaction; ss		
OS	Synthetic.		
XX			
FT	Key	Location/Qualifiers	
FT	misc_feature	1..1219	
FT		/tag= a	
FT		/note= "NheI/SpeI kan cassette"	
FT	CDS	121..936	
FT		/tag= b	
FT		/note= "kan gene"	
FT	misc_feature	1220..2011	
FT		/tag= c	
FT		/note= "E.rep cassette"	
FT	misc_feature	2012..3920	
FT		/tag= d	
FT		/note= "M.rep cassette"	
FT	misc_feature	3925..4120	
FT		/tag= e	
FT		/note= "synthetic multiple cloning site"	
XX			
PN	US5583038-A.		
XX			
PD	10-DEC-1996.		
XX			
PE	21-OCT-1991;	91US-0780261.	
XX			
PR	17-NOV-1992;	92US-0977630.	
PR	21-OCT-1991;	91US-0780261.	
XX			
PA	(MEDT-) MEDIMMUNE INC.		
XX			
PI	Stover CK;		
XX			
DR	WPI; 1997-042315/04.		
XX			
PT	Mycobacteria expressing secretion signal of lipoprotein and		
PT	heterologous antigen, esp. outer surface protein A or B of Borrelia		
PT	burgdorferi - are used in the form of a live bacterial vaccines		
PT	against Lyme disease		
XX			
PS	Example 1; Fig 12; 122pp; English.		
XX			
CC	This sequence represents the plasmid pMW206. This plasmid was		
CC	used in the construction of a series of vectors used to produce		
CC	the recombinant Mycobacteria of the invention. The recombinant		
CC	mycobacteria of the invention are transformed with DNA encoding		
CC	a polypeptide which comprises a lipoprotein secretion signal		
CC	sequence and an antigen (Ag) heterologous to the mycobacteria.		
CC	The lipoprotein secretion signal causes the Ag to be produced as		
CC	a lipoprotein. The mycobacteria may be used in the form of a live		
CC	bacterial vaccines against Lyme disease, where the bacteria express		
CC	a surface protein of Borrelia burgdorferi, the causative agent of		
CC	Lyme disease.		
XX			
Sequence	4120 BP; 922 A; 1125 C; 1166 G; 907 T; 0 other;		

[illegible]



PA (PATH-) PATHOGENESIS CORP.  
XX  
XX Hickey MJ, Stover K;  
XX WPI: 1996-209365/21.  
DR  
XX  
XX  
PT Quantification of mycobacteria by luminescence - using bacterial  
PT reporter strains contg. a lux gene, permits the determ. of the  
PT efficacy of anti-mycobacterial and prophylactic compsns.  
XX  
XX  
PS Example 1: Page 58-60; 82pp; English.  
XX  
XX This is the complete sequence of plasmid PMH28. This plasmid has  
CC been constructed by inserting a multiple cloning site and an  
CC Escherichia coli T172 terminator sequence into plasmid pPA207SC,  
CC based on plasmid PMV206 (AAT28281). A multiple cloning site is  
CC inserted into PMH28 to give plasmid PMH29 (AAT28273), and a mutant  
CC firefly luciferase (EC-1.13.12.7) reporter gene is inserted to give  
CC plasmid PMV30-lux (AAT28278). This type of vector may be used in a  
CC new method for quantifying Mycobacterium infection in vivo. An  
CC animal is injected with e.g. Mycobacterium bovis BCG, etc.,  
CC transfected with one of these vectors, and reporter gene expression  
CC is detected by luminescence in tissues without lysis or cell  
CC concentration. The method may be used to determine the efficacy of  
CC antibiotic, tuberculostatic and vaccine compositions.  
XX  
SQ Sequence 4236 BP; 962 A; 1181 C; 1204 G; 949 T; 0 other;  
  
Query Match 2.3% Score 80.6; DB 17; Length 4236;  
Best Local Similarity 47.6% Pred. NO. 1.4e-09;  
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;  
  
QY 301 agaagctgctgcacagcagcgtgctgacgagcaaaatctgggctgacggcagc 360  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2405 agcctgtgctgcgtacgtgcgcgtgcaagcagatctctgaggggtctacggcc 2464  
QY 361 tgaactggcagaggcgctgagctgacctatcagacgaaccgctgtgcatgcagt 420  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2465 aaagcgcgcgtgcgcctcagcgcgcgtatcagcagaccacacagcgtcgcaa 2524  
QY 421 ccttggtcatcacccgatgatgcttgcgtgctgactggtggccgacgtcgtgggc 480  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2525 accgtgcgtgtgtgagctagacatccacagcagcgctcagcgctcccgccg 2584  
QY 481 tgcctcacccgtctcagctgtcctatg-----aaccggtcacgacccacgacacatcg 534  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2585 gttccatcgcgtcccaacgcgctgtgtggcaatcgcgcaacggccagcagcagcag 2644  
QY 535 tctatgccttgaagaacctgtgtctgacgagatgcgcgcgcgacgacctataacc 594  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2645 tgttggaactcaaacgcccttctcaacgacgagatacgcgcgctgtaagcgcgtcgat 2704  
QY 595 tgcgcgcgcggtgtagagaggtctatgtagcgtcttcgscggtcgatgatacctcagggc 654  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2705 acaatggcgcgctgtagcgaagccttcggtcgcgcgctgtagcgaacgtagtactcag 2764  
QY 655 accggtatcaaaaagacccggtcagcaccgacatgcaacctctggggcccgacagc 714  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2765 gccatcagacaaaaccccgccgacatcgctgtggaaacggaatgtgtcattcagatc 2824  
QY 715 cgtctacgagcttgcgcgcctgcacacacccctgacgagatcaccgactgcggaag 774  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2825 tctacacatcagcacatcagagcgtcgctgcgcgacacatgcacacgcgcgtggc 2884  
QY 775 cagggaaccgcgtcgcgaacgtcacccgatcaaggttcgscgcaagcgtcacctgttcg 834  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2885 gtacgagacacagctlaaaaagcgtctcgacgacgctagggcgtgaattgcactgttcg 2944  
QY 835 acacacccgcgactgggacataccagggccgtcc 867  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2945 attcgcacaggttctgggctatcgtccgcgc 2977

RESULT 8  
AAT28273  
ID AAT28273 standard; DNA: 4352 BP.  
XX  
XX  
AC AAT28273;  
XX  
XX 19-AUG-1996 (first entry)  
XX  
XX  
DE Plasmid PMH29 sequence.  
XX  
XX  
KW Plasmid PMH29; vector; firefly; luciferase; DNA cassette;  
KW reporter gene; luminescence; BCG; Mycobacterium bovis;  
KW Mycobacterium tuberculosis; Mycobacterium avium;  
KW Mycobacterium intracellulare; Mycobacterium kansasii;  
KW Mycobacterium scrofulaceum; Mycobacterium leprae; tuberculosis;  
KW leprosy; antibiotic; tuberculostatic; vaccine; drug screening; ss.  
XX  
XX Synthetic.  
XX  
XX  
XX WO9610645-A1.  
XX  
XX 11-APR-1996.  
XX  
XX 02-OCT-1995; 95WO-US12642.  
XX  
XX 03-OCT-1994; 94US-0316950.  
XX  
XX  
PA (PATH-) PATHOGENESIS CORP.  
XX  
XX Hickey MJ, Stover K;  
XX WPI: 1996-209365/21.  
DR  
XX  
XX  
PT Quantification of mycobacteria by luminescence - using bacterial  
PT reporter strains contg. a lux gene, permits the determ. of the  
PT efficacy of anti-mycobacterial and prophylactic compsns.  
XX  
XX  
PS Example 1: Page 61-63; 82pp; English.  
XX  
XX This is the complete sequence of plasmid PMH29. This plasmid has  
CC been constructed by inserting a multiple cloning site and an  
CC Escherichia coli T172 terminator sequence into plasmid pPA207SC,  
CC based on plasmid PMV206 (AAT28281), to give plasmid PMH28 (AAT28272)  
CC and insertion of a multiple cloning site to give PMH29. A mutant  
CC firefly luciferase (EC-1.13.12.7) reporter gene is inserted to give  
CC plasmid PMV30-lux (AAT28278). This type of vector may be used in a  
CC new method for quantifying Mycobacterium infection in vivo. An  
CC animal is injected with e.g. Mycobacterium bovis BCG, etc.,  
CC transfected with one of these vectors, and reporter gene expression  
CC is detected by luminescence in tissues without lysis or cell  
CC concentration. The method may be used to determine the efficacy of  
CC antibiotic, tuberculostatic and vaccine compositions.  
XX  
XX  
SQ Sequence 4352 BP; 979 A; 1196 C; 1217 G; 960 T; 0 other;  
  
Query Match 2.3% Score 80.6; DB 17; Length 4352;  
Best Local Similarity 47.6% Pred. NO. 1.4e-09;  
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;  
  
QY 301 agaagctgctgcacagcagcgtgctgacgagcaaaatctgggctgacggcagc 360  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2405 agcctgtgctgcgtacgtgcgcgtgcaagcagatctctgaggggtctacggcc 2464  
QY 361 tgaactggcagaggcgctgagctgacctatcagacgaaccgctgtgcatgcagt 420  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2465 aaagcgcgcgtgcgcctcagcgcgcgtatcagcagaccacacagcgtcgcaa 2524  
QY 421 ccttggtcatcacccgatgatgcttgcgtgctgactggtggccgacgtcgtgggc 480  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2525 accgtgcgtgtgtgagctagacatccacagcagcgtcgcgctcagcgcccgccg 2584







	Query Match	2.3%	Score 80.6;	DB 17;	Length 6171;
	Best Local Similarity	47.6%;	Pred. No. 1.5e-09;		
	Matches 273;	Conservative	0;	Misses 294;	Indels 6; Gaps 14.
OY	361 agagctggtcccaacgaacgcgtcggcgctcaaggccaagaatcttgggctaccggcacg	360			
Db	2382 agctcttgcttcgcgtactcgtgcgcgcctggacaagcgaacgatctgtcgaaggatctaacgcc	2441			
OY	361 tgaactcgcagaaaggcgctbgagactctgcctatacctgaagcgaacctgtgtcatgtcagt	420			
Db	2442 aaagcgcgcgttcgcgccttaaggccgcgglatacatcgaagcgaaccaacagcgtctgcaa	2501			
OY	421 ccttgtatcataccgatcgagatgttttgatgtcgtacgtggccgcgcagaccctcgtctggc	480			
Db	2502 aacctgtctgtctgtgagctagtaacatacgaacgcagcgcctccgaagcgcgtacgcgcggg	2561			
OY	481 tgccttcacgcgtctcactacgtgtctcag-----aacccgtgtcacgaacaccggaacacatcg	534			
Db	2562 ggtcccattcgcgttcgccaacgcgcatcgttgggaatctgcgcgaacgcgcacgaacacgcag	2621			
OY	535 tctatgtccttgaagaacctgtgtgtctgtacccgatgtccgcgcgcgcgcgcgtctataacc	594			
Db	2622 tgttggcaactcaacgcctctgttccacgaacccgaatacgcgcgcgcgttagaccgcgtcgcac	2681			
OY	595 tgctgcgccgcgtcgcgaagcgcgcgttatgtcgagcttctcggcgcgagatgcatactacgggc	654			
Db	2682 acatgtgcgcgcgtgtgcgcgaagcctcttcggcgcgccgttgatgtgcgcgcgcgattactcag	2741			
OY	655 accggatacaaaaagaaccgcgtcgaacgcgcgcacatgtgcacactctctgggccccgcgaagcg	714			
Db	2742 gccctatgacccaanaaacccgcgcacatcgcctctgggaacggatgtgtccactcagatic	2801			
OY	715 cgctctacgaagctgtgcgcgcctcgcacaacacctcgaagagatctcaacgcacgtgcggaag	774			
Db	2802 tctacaactcagccacacatcgaagcgcgaagctcgcgcgaacatgcacacgcgcgcgttgc	2861			
OY	775 caagggaaccgcgtctgcgaacgtcacacgcgtacacaggtctgcgcgcgcgaacgttacacctgtcg	834			
Db	2862 gtccagcagacacgatacaaaagcgcgtctcgaagcgcgttagtcggaattgcacattctcg	2921			
OY	835 aacacaccgcgcatgtgtgcgtataccggccgctcc	867			
Db	2922 attccgtcaggtgtgtggcctatcgttccgcgc	2954			
	RESULT 12				
ID	AAT28281				
	AAT28281 standard; DNA; 4120 BP.				
AC	AAT28281;				
XX					
DT	19-AUG-1996 (first entry)				
XX					
DE	Plasmid pMW206 sequence.				
XX					
KW	Plasmid pMW206; vector; firefly; luciferase; DNA cassette;				
KM	reporter gene; luminescence; BCG; Mycobacterium bovis;				
KM	Mycobacterium tuberculosis; Mycobacterium avium;				
KM	Mycobacterium intracellulare; Mycobacterium kansasii;				
KM	Mycobacterium scrofulaceum; Mycobacterium lepre; tuberculosis;				
KW	leprosy; antibiotic; tuberculostatic; vaccine; drug screening; ss.				
XX					
OS	Synthetic.				
XX					
PN	WO9610645-A1.				
XX					
PD	11-APR-1996.				
XX					
PF	02-OCT-1995; 95MO-US12642.				
XX					
PR	03-OCT-1994; 94US-0316950.				

[illegible]



PR 21-OCT-1991: 91US-0780261.  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 XX  
 PI stover CK;  
 XX  
 XX  
 DR WPI: 1993-152187/18.  
 DR P-PSDB; AAR34542; AAR34543; AAR34544.  
 XX  
 XX  
 PT Expression vector for expressing protein or polypeptide in  
 PT mycobacterium - contg DNA sequences encoding lipoprotein  
 PT secretion signal and peptide heterologous to bacteria expressing  
 PT fusion protein of lipoprotein heterologous to bacteria  
 PS  
 PS Example 1; Fig 5; 86pp; English.

CC This sequence represents that of plasmid pMV101. This is a modified  
 CC version of plasmid pYUB125, which had 792 bases of the tet gene  
 CC (inactivated by prior manipulations) deleted. The plasmid  
 CC represents an expression vector. The HSP61 coding sequence was  
 CC inserted between the NheI and BamHI sites to form pMV65A.  
 CC NOTE: - When the proteins encoded by this sequence were decoded,  
 CC the published sequence was found to contain approx. 65 codons which  
 CC did not code for the amino acids dictated by the universal genetic  
 CC code. The sequence also contained many insertions and deletions  
 CC which upset the open reading frame that was displayed as the  
 CC encoded proteins. The complementary strand was also found to  
 CC contain many insertions and deletions that did not have corresponding  
 CC complementary bases on the coding strand.

XX Sequence 6407 BP; 1405 A; 1892 C; 1670 G; 1439 T; 1 other;

Query Match 1.9%; Score 68.6; DB 14; Length 6407;  
 Best Local Similarity 47.5%; Pred. No. 1.3e-06;  
 Matches 272; Conservative 0; Mismatches 294; Indels 7; Gaps 2;

QY 301 aagagctggtcgcagcagccgctggtcagcagcagagcttggtggtacagcagc 360  
 DB 2853 AGCTGTGGCTGCCGTACTGGGGCTGCAAGCAGCATCTG-CTCAGAGGGGATCTAGCC 2795  
 QY 361 tgactcgagagagggcgctgagctgcttacatcgaaagccgctgtgtatcagt 420  
 DB 2794 AAGGCGCGCGTGGCGCCGAGGCGCGCGTACATCGAGGGAACCAAGAGGGCTGGCA 2735  
 QY 421 ccttggtacacagctgagatgctgagatgctgagctgagcagcagcctgcctgggc 480  
 DB 2734 ACCTGCTGCTGTGAGCAGTACCATCCAGACGACGCTCCAGAGGCTTAGCGCGGG 2675  
 QY 481 tgccttcacgctcctactgctcat-----gaaccgtgtcagacacagcagacatcg 534  
 DB 2674 GGTCCCATCCGCTGCCCAACGCGATGCTGGGCAATGGCGCAACGGCCAGCAGCAG 2615  
 QY 535 tctatgcttgaagaacctgtgtgtctgtacgagtcgcgagcgagcgactataacc 584  
 DB 2614 TGTGGGGAAGTCAAGCGCCCTGTCTCCAGCAGCAGCAATACGCGGCTTAGAGCCCTG 2555  
 QY 595 tgcgcgcgcgcgtcgagcagcgagcctatgagcgttctcgagcgagcagatgactacgggc 654  
 DB 2554 ACATGCGCGCGTGGCCGGAAGGCTTCGCGGCGGCTGCAAGGCGGACCCAGTTACTCAG 2495  
 QY 655 accgagcaaaagaacccgctcgaacacccatgcaacccctctggggcccccagagcg 714  
 DB 2494 GCCTCATGACCAAAAAGCGCGGCGCATCGCTGGGAAAGGAAATGGTCTACTCAGATC 2435  
 QY 715 cgctctagagctggtgcgctcctgcacacacacctgagagatccacgcaatcgcgagcg 774  
 DB 2434 TCTACACACTGACGACATCTGAGCGAGCTGGGGCGGAACCAACGCCGCCGCTGGCG 2375  
 QY 775 cagggagaccgcgctcgacagctcaccgatacagcgctcgagcgcaacgctacacctgtcg 834  
 DB 2374 GTGAGCAGACCAAGTCAAAAGCGGCTCCGACGCCGCTAGGGCGGAATTTGGCAGCTGTCG 2315

QY 835 acaccacccgcatgtggcataccggcgctcc 867  
 DB 2314 ATTCGCTCAGGTTGTGGCCTATCTCCGGCC 2282

RESULT 15  
 AAQ31743/C  
 ID AAQ31743 standard; DNA; 4811 BP.  
 XX  
 XX  
 AC AAQ31743;  
 XX  
 XX 18-FEB-1999 (first entry)  
 DT  
 XX  
 DE PMV101 - a deletion mutant of pYUB125.  
 XX

KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;  
 KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;  
 KW pertussis; malaria; influenza virus; CTL; herpes virus; ds.  
 XX  
 XX Mycobacterium.

FT Key Location/Qualifiers  
 FT misc\_difference 2703..2705

FT /tag= a  
 FT /note= "This codon represents 1700 bases missing  
 FT from the sequence reproduced in the  
 FT specification"

FT cds complement (1598..2545)

FT /tag= D  
 FT /product= tet e protein

FT /tag= C  
 FT /product= complement (1956..2700)

FT /tag= C  
 FT /product= tet d protein

FT /note= "precise start posn. unknown as it is  
 FT contd. in missing portion of sequence."

FT cds 3661..4474

FT /tag= d  
 FT /product= kan protein

PN M09221374-A.

PD 10-DEC-1992.

PF 01-JUN-1992; 92MO-US05023.

PR 06-JUN-1991; 91US-0711084.

XX (MEDI-) MEDIMUNE INC.

XX Dela Cruz V, Stover CK;

XX WPI: 1992-433378/52.

XX Tetanus vaccination - by provoking an immune response using

XX transformed Mycobacteria

XX Example 1; Fig 5; 86pp; English.

XX This sequence represents that of plasmid pMV101. This is a modified

XX version of plasmid pYUB125, which had 792 bases of the tet gene

XX (inactivated by prior manipulations) deleted. The plasmid

XX represents an expression vector. The HSP61 coding sequence was

XX inserted between the NheI and BamHI sites to form pMV65A.

XX NOTE: - When the proteins encoded by this sequence were decoded,

XX the published sequence was found to contain approx. 65 codons which

XX did not code for the amino acids dictated by the universal genetic

XX code.

XX Sequence 4811 BP; 1076 A; 1376 C; 1225 G; 1130 T; 4 other;

Query Match 1.6%; Score 58.6; DB 13; Length 4811;  
 Best Local Similarity 47.6%; Pred. No. 0.00032;

	Matches	172:	Conservative	0:	Mismatches	189:	Indels	0:	Gaps	0:
QY	507	aaccggtc	acgacacacgacacatcgtctatgcttgaagaaacctgtgtgtcgtgacc	566						
Db	2642	AATCGCGCAACGGCGACACGACAGTGTGGGACTCAACGCCCTGTTCACAGCAC	2583							
QY	567	gattccgagcgagcgagccatacaacctgtgtccggtcagagcgagccatgtgac	626							
Db	2582	GAATACGCGCGCGCTAAGCCGCTCGCATACATGGCGGCGTGGCGAAGGCCCTTCGGCGC	2523							
QY	627	gtctcggcgagtgatcctacgagcgacggtacaaagaaaccgctcagcacgagc	686							
Db	2522	GGCGTCAAGCGCGACCGAGTTACTCAGGCTCTCATGACCAAAAAGCCGGCCACATCGCC	2463							
QY	687	catcgacacctctgtgggcccccgacgagcgctctacgagctgtcgccccctgcacacacc	746							
Db	2462	TGGGAACGGAATGGGTCCACTCTACACACTCAGCCACATCGAGCGGAGCTG	2403							
QY	747	ctcgacgagatccacgacatgcgagagcgaggaaccggtcgcgaagtcacccgatca	806							
Db	2402	GGGGCGAACAATGCCACCGCCGCTGGGTGAGCAGACCAAGTACAAAGCGGCTCCGACG	2343							
QY	807	acggtcggcgcaacgtacacctgttcgacacacccgcatgtgagcataccgggagcgtc	866							
Db	2342	CCGCTAGGGCGGAATGTGGCACACTGTTCGATTCGCTCAGGTTGTGGGCTATGTCGGGCC	2283							
QY	867	c 867								
Db	2282	C 2282								

Search completed: October 6, 2002, 16:32:33  
 Job time: 18386 sec









and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpbes@sc.riken.go.jp, URL: http://gsp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PBH. This BAC end  
was generated during the Rad process and may have higher chance of  
clone tracking errors.

## PRIMERS

Sequencing: M13Rev

## LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 1091

/organism="Pan troglodytes"

/db.xref="taxon:9598"

/clone="PBH-089D22.R"

/sex="male"

/cell\_type="lymphoblast"

/clone\_id="PBH Chimpanzee Male BAC Library"

30 a 307 c 601 g 152 t 1 others

## BASE COUNT

30 a 307 c 601 g 152 t 1 others

## ORIGIN

Query Match 1.5%; Score 55; DB 12; Length 1091;

Best Local Similarity 45.6%; Pred. No. 0.74;

Matches 193; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

## FEATURES

source 1. 874  
Location/Qualifiers

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="Plate=1081 Col=6 Row=1"

/clone\_id="RP11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

143 a 183 c 412 g 86 t 50 others

## BASE COUNT

143 a 183 c 412 g 86 t 50 others

## ORIGIN

Query Match 1.5%; Score 53; DB 12; Length 874;

Best Local Similarity 47.9%; Pred. No. 1.7;

Matches 146; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

## FEATURES

source 1. 874  
Location/Qualifiers

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="Plate=1081 Col=6 Row=1"

/clone\_id="RP11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## BASE COUNT

143 a 183 c 412 g 86 t 50 others

## ORIGIN

Query Match 1.5%; Score 53; DB 12; Length 874;

Best Local Similarity 47.9%; Pred. No. 1.7;

Matches 146; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

## FEATURES

source 1. 874  
Location/Qualifiers

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="Plate=1081 Col=6 Row=1"

/clone\_id="RP11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## BASE COUNT

143 a 183 c 412 g 86 t 50 others

## ORIGIN

RESULT 6  
A0744715/c 874 bp DNA linear GSS 16-JUN-1999  
LOCUS HS\_5505\_A2\_E03.SP6.RP11-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate=1081 Col=6 Row=1, DNA sequence.  
ACCESSION A0744715  
VERSION A0744715.1 GI:5522237  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
1 (bases 1 to 874)  
Matsuzaki, S., Matsuzaki, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L. Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L. Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Matsuzaki, S., Wallace, J.C., Hood, L.  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RP11-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 1081 row: 1 column: 6  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 874.  
Location/Qualifiers

LOCUS	B96799		570 bp	DNA	linear	GSS 31-MAR-1998
DEFINITION	T32L9R TAMU Arabidopsis thaliana genomic clone T32L9,					DNA sequence.
ACCESSION	B96799					
VERSION	B96799.1					GI:2998045
KEYWORDS	GSS.					
SOURCE	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.s.
REFERENCE	1 (bases 1 to 570) Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K, Berry,R.K., Granger,D., Suh,E., Wilde.C., Adams,M.D. and Venter					,D.C.
AUTHORS	A BAC End Sequence Database for Identifying Minimal Overlaps In Arabidopsis Genomic Sequencing. Update 3 Unpublished (1997)					
TITLE	JOURNAL					
COMMENT	Contact: Steve Rounsley Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0200 Fax: 301 838 0208 Email: rounsley@tigr.org Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 570. Location/Qualifiers  1..570 .organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="T32L9" /clone_lib="TRAMU" /sex="thermaphrodite" /note="Vector: BelobACIT; Site_1: HindIII; Site_2: HindIIII					
FEATURES	SOURCE					
BASE COUNT	ORIGIN					
	Query Match                1.5%, Score 52.6; DB 12; Length 570; Best Local Similarity     45.5%; Pred. No. 1.9; Matches 187; Conservative      0; Mismatches 224; Indels          0; Gaps            0;					
OY	578 ggcagcgctatacaactcgtctccgcytcggatcgagaaggacctatgccgaagtcttgccg					
Db	4 gcTTCCGGCGCTTGCGGAGGCCGC CGGGGAATACC GCCCGCGTCGGCGMACACTTGCTCA					63
OY	638 cgatgatcatctacagcggaccgatatcaaaaagaaccgctcacgcgccatcgcgcct					697
Db	64 CAATAACGCCCAAAGCCAAGGTCTGCGCCCCGCCGCTGCTCTCCGCCGAGA GCCCGTAGCC					123
OY	698 ctggggcccgcgaagcgcgtctacaagctcgcgccccctcgcgacacacacctcgaagat					757
Db	124 CTTGCTCGCGCCC GTGGCCGCGAAGCTCGTAAC TCGAAC CCGAAAC ACCCTCACCGCGC					183
OY	758 caacgaattgcggaggcaggaaaccgcgctcgcgaactgaacgttaacccgatacagytgcgcg					817
Db	184 CGTGGCTGTGTGCGGTGGCGAGATCATCGAGGCC CCACGCGCATCTGANTGCCGATGGCAC					243
OY	818 caacgtaaccttgtttcacaccaaccgcgatgtgggcataaccggcgcttcggcatactcgy					877
Db	244 CGGCTTCAGCGAGGTGTCGTCCGATTGGCTTCGCTCCGACACC AGGCCGCCACAGCGCCA					303
OY	878 ggsgcggccggttcgcggaatggagacacacgltatcggagacataccaactctaagca					937
Db	304 GAACAATCTACAGCGCCGATNGTGCGCCAGCA CGTCGCCGATGTGAGCGCGACTTGACTCT					363
OY	938 gaagtcatacgccgaagaattgcacaaaggcccccttggtttaacgaact					988

DB	364	CGATACCCCTGCACACGCTGCGGAACCGACCCGCGCAACATCTCT	414
RESULT	8		
LOCUS	AM12314		
DEFINITION		448 bp mRNA linear EST 30-NOV-2001	
ACCESSION	AM12314		
VERSION	AM12314.1		
KEYWORDS	EST.		
SOURCE	soybean.		
ORGANISM	Glycine max		
REFERENCE			
AUTHORS	Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Corvett, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: cdu@resgen.com Seq primer: -40RP from Gldco High quality sequence stop: 401. Location/Qualifiers 1. 448 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl013-2355" /clone_1db="Gm-cl013" /tissue_type="Whole seedlings, 2-3 week old seedlings, greenhouse grown" /lab_host="XLI0-Gold" /note="Vector: Bluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into XLI0-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg."		
BASE COUNT	84 a	154 c	151 g
ORIGIN			59 t
Query Match	1.44;	Score 51.4;	DB 9; Length 448;
Best Local Similarity	47.44;	Pred. No. 3;	
Matches 154;	Conservative 0;	Mismatches 171;	Indels 0; Gaps 0;
653	gcaccgatacaagaacccgcgtatgacgacgcccacatcgacctctctgggcccgcga	712	
73	ggaccgcgacacgaagacacgctgcggctgacgctccgacgctcgcacgacgacg	132	

QY	713	cgcgcgtctacgagctcgcgcgccttcgcaacacaccccttcagcagatccacgacactgcgga	772
Db	133	CGGCTTGGGGCCAGAGAGACTGCTGTGGCAACGCCACCAATGCTTTAAGCCCTTGGCGCA	192
QY	773	ggcagaggaaaccgcgcgcgaacgltcacccgatlcaacgyltcgycgcgaacgyltaccctglt	832
Db	193	CTCGAGGCACACTAGTACGCCGCCAGGTGTCCGACGACCTGTACTCAGCGGGCTCGCA	252
QY	833	cgacacccaccgcgcggtggcgtacccgycgcgtccgcgactctctggcggcgccggtgc	892
Db	253	CGGCAACGACACCGACTGGACTTCTGTGGAGGGTACGGCGACCGGCTTCGCCGG	312
QY	893	cgaaatgggaagacacggtatctcgagacacatccactactgaaacggaacgacatcatcgca	952
Db	313	CGAGCGCAGCCCTCATTCTGTGTGGGCGCTCGGCTGCACCGGGGACGAGGGCAACTTACGA	372
QY	953	cgaaatcgcaacagggccccctcgcg	977
Db	373	GTAACCTGGCAACTGCTCAACGGC	397

RESULT	9
LOCUS	BI144217
DEFINITION	BI144217 1260 bp mRNA linear EST 05-JUL-2001
ACCESSION	602907950F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5064900
VERSION	5', mRNA sequence.
KEYWORDS	BI144217 BI144217.1 GI:14604218
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 1260)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: c9abbs-r@mail.nih.gov  
 Tissue Procurement: jeffrey E. Green, M.D.  
 cDNA library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM1175 row: c column: 13  
 High quality sequence set: 405.

FEATURES	Location/Qualifiers
source	1. .1260

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5064900"
/clone_1lb="NCL.CGAP_Kid14"
/lab_host="DH10B (p1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCL.CGAP Library. |"

```

Query Match	1.48;	Score 51.4;	DB 10;	Length 1260;
Best Local Similarity	47.48;	Pred. No. 4;		
Matches 154;	Conservative 0;	Mismatches 171;	Indels 0;	Gaps 0;

OY 571 ccgcgcgagcgaacggctatacaacctgtctgccgcgctgcagcaggcctaagcagcttc 630

Db 628 cccatgcccccacgagcccccccccttccaccacccacacccccccgacgctgcggccccc 687

[illegible]

RESULT	10
CNS01087	
LOCUS	CNS01087
DEFINITION	CNS01087 Drosophila melanogaster genome survey sequence Sp6 end of BAC BAONC03H07 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL098641
VERSION	AL098641.1 GI:5610252
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscumorphia; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 961)
REFERENCE	

AUTHORS	TITLE	JOURNAL	COMMENT
Genoscope. Direct Submission	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:segre@genoscope.cns.fr">segre@genoscope.cns.fr</a> )		Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CNRS (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES	Location/Qualifiers
source	1. .961

```

/organism="Drosophila melanogaster"
/plasmid="pBelosBAC11"
/db_xref="taxon:7227"
/clone_1fb="DrosBAC"
/clone="BACN03H07"
/note="end : Sp6"

```

Query Match	1.48;	Score 51.2;	DB 12;	Length 961;
Best Local Similarly	25.58;	Pred. No. 4.1;		
Best 69; Conservative	91;	Mismatches 111;	Indels 0;	Gaps 0;

[illegible]

QY 635 cggcgatgcattcctacgcygcacggaatcacaaagccgtcagcaaccgcccatgagac 694  
|| : : : || : : : || : : : | : ||||| || : : |  
Db 738 CGMAAATAAAVMMARRGVCAMCKCASDSASAVCMHGGCCGSCACCCCCSCSCSGC 797

[illegible]

FEATURES	SOURCE
BASE COUNT	257 a 170 c 162 g 96 t 250 others
ORIGIN	

[illegible]

FEATURES	found at <a href="http://bacpac.med.buflalo.edu/drosophila_bac.htm">http://bacpac.med.buflalo.edu/drosophila_bac.htm</a> .
source	Location/Qualifiers 1. .932 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BAC14B09" /note="end : 77"
BASE COUNT	155 a 202 c 241 g 91 t 243 others
ORIGIN	
Query Match	1.4%, Score 50.6; DB 12; Length 932;
Best Local Similarity	28.1%; Pred. No. 5.3;
Matches 133; Conservative 104; Mismatches 236; Indels 0; Gaps 0;	
QY	601 cccgcgtcgagcgaagcgtctatgcaacgtcttctggcgcgatcatcctacgagcaccga 660
DB	922 CCGSGGCGSSSSCGGSGCGSSGCGSSGCGSCCGCGSSCGSCCGCGCGCGCGGCGGCGSGCG 863
QY	661 tcacaagaaccgcgttcagccacgcacatgcaacctctggggcccgacagcgctct 720
DB	862 CCGCGGSSSSCG 803
QY	721 acgagctgcgcgccttcgcacacaccccttcgaagatccacgacgtcgcgagagga 780
DB	802 SCSSCGCGSSCG 743
QY	781 accgcgcgttcgaacgtcaccgataaaggtctgcgcgcaagtcacccctgttcgacaca 840
DB	742 GSGGCGSGSGGSGCGSGCGSGCG 683
QY	841 ccgcgatgtggcgtacacagcgcgctgcgcgcacatccttggcgcgccggttcgcgaatgg 900
DB	682 SSGCGCGCGSGCG 623
QY	901 agcacaccgtatctgcagacatcacactactgaacgagacgatacatcgcgcgaatgg 960
DB	622 GSGCG 563
QY	961 ccacagcgcccttcggttgaaacgaacttaagcattatcgtatccatttcgcgatgg 1020
DB	562 CCSCMSGCCCG 503
QY	1021 tctggcgaacttcaccccccgaaccttcgcgcgcgcgcgcgaagaacgatacgc 1073
DB	502 CCG 450
RESULT 14	
AV623353	498 bp mRNA linear EST 15-DEC-2000
LOCUS	AV623353 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION	reinhardtii cDNA clone LC062e06_r 5', mRNA sequence.
ACCESSION	AV623353
VERSION	AV623353.1 GI:10772530
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii.
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
AUTHORS	Chlamydomonadaceae; Chlamydomonas.
	1 (bases 1 to 498)
	Asamiizu, F., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohnaya, K.,
	Nakamura, Y. and Tabata, S.
	Generation of expressed sequence tags from low-CO2 and high-CO2
	adapted cells of Chlamydomonas reinhardtii
	DNA Res. 7 (5), 305-307 (2000)
	20539644
TITLE	Contact: Erika Asamiizu
JOURNAL	The First Laboratory for Plant Gene Research
MEDLINE	Kazusa DNA Research Institute
COMMENT	Kazusa 1533-3, Kisarazu, Chiba 292-0812, Japan
	Email:asamiizu@kazusa.or.jp URL: <a href="http://www.kazusa.or.jp/en/plant/">http://www.kazusa.or.jp/en/plant/</a>

FEATURES	
SOURCE	Location/Qualifiers 1..498 /organism="Chlamydomonas reinhardtii" /strain="Cg" /db_xref="taxon:3055" /clone="LC62ae06_r" /clone.lib="Chlamydomonas reinhardtii 5% to 0.04% CO <sub>2</sub> " /note="Vector: pluscriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"
BASE COUNT	90 a 184 c 146 g 78 t
ORIGIN	
Query Match	1.4%; Score 50.4; DB 9; Length 498;
Best Local Similarity	46.9%; Pred. No. 4.9;
Matches 196; Conservative	0; Mismatches 216; Indels 6; Gaps 1;
OY 493	cctacgtatcattgaaccggtltaacagccaccggacacatcgltatgccttgaagaacc 552
36	CCCACAAAGTGCAGAACAATGGTGTTCCACGACAGCTGCACCGTGAACGCCAATGCCCATG 95
OY 553	cttgtgtcttgaccggatgccgcggcgaggagcgtcatcaactgttcgccgctcgagc 612
Db 96	AAGTGCACCATCACCGACAGTCGCCGCGCTCAAGAGCCATCTCCCCACGCGCCACTTAAGC 155
OY 613	aggagccatagcagcgtcttcggcgagtagtatccctacagggcacccggatccaagaacc 672
Db 156	TGTTCACAGGAGGACATGTTCCGCGACCCTGGCCGTTCTCCGACAAGACATGGCCCATCTGCC 215
OY 673	cgctcagcagccgcatlbgcagacctlbtggyggcccgcagagcgcgtctcagctgcg 732
216	TGCTGCGCTACTTCAACCCCAATCGCGGCGCAACCCAGGGGAGAGTGGGTAGACACCCG 275
OY 733	ccctcgacaacacctcgagagatcaacagcagtgcggaggacaggaaaccgcgctgca 792
Db 276	TGGGCATCCCAACAACCTCATGCCGTACATCCAGCAAGTAGTGCTGGGGCCAAGCG----- 330
OY 793	acgtcacccgatcaacggttcggcgcgcaacgtcaccctlttcgaacacaccccgcatgtgg 852
Db 331	-CGAGTTCCTGCGGTGTGTTGGCAACGACTACCCACGCCCGACGGCACCGCATCTCCCG 389
OY 853	cataccggagccgctccggagactccttggggagcccgccgyltcgcgaatgaggagacacgct 910
Db 390	ACTACATCATCAGTGTATGAGCTGGCGGAGGGGCACAGTCAGGCGTGTGTCAAAGACCT 447
RESULT 15	
CNS01213	
LOCUS	645 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC
VERSION	BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
ACCESSION	AL101589
VERSION	ALI01589.1 GI:5613200
SOURCE	GSS.
ORGANISM	fruit fly. Drosophila melanogaster Euarystota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 645) Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CDPH (Centre
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

```
d'Eitude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.
```

```
FEATURES
    source                location/Qualifiers
                        1..645
                        /organism="Drosophila melanogaster"
                        /plasmid="pBelOBAC11"
                        /db_xref="taxon:7227"
                        /clone_1ld="DrosBAC"
                        /clone="BACN08C07"
                        /note="end : T7"

BASE COUNT      28 a          26 c          85 g          92 t         414 others
ORIGIN
```

```
Query Match              1.4%; Score 50; DB 12; Length 645;
Best Local Similarity   9.5%; Pred. No. 6.3;
Matches 40; Conservative 131; Mismatches 249; Indels 0; Gaps 0;
```

```
DQ 293 gtccctgagacgtgcgtcaccagcaaacgcgtggcgctcacgacgaagtctggggcga 352
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 GTTNNNGNNNNNNSCCCSNCTNCNNNTNTNTSSSSSSNNNNNTNNNSNNNSNN 236

QY 353 gcgcgacgtgactgcgcgacagaaggcgctgtagcttcattacatcgaaacccgttgt 412
::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 237 NNSSSGSSNNNNNGGSSSSNNNSNNGSGSSSSSNSNTSSNNNSNNNTSSNNNSNN 296

QY 413 catcgactccttgctcatcacacgatcgatgcttcgagtctgactggccgcagacct 472
::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 297 TNSNNNSNNSSSTTTNTSNTSSNNSATGSSSSNSVSGTTBSGSSSSSSNNNNN 356

QY 473 cgctggcgctgccttccactctactagttcatagAACCGTGTCAAGACCAGCACCAT 532
::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 357 NNANNSBTSNNNSNNNSMATTSTINSSSSSSSSSSSTSSNTSTSTSSNSMTSTT 416

QY 533 cgtctatgacctgaagAACCGTGTGTCTAGCACGATGCCGCgcgcgacgacctataa 592
::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 417 SSSSTTNSTSTSSSSNSTGSSSSTTTTNSASNNNNNNNNNGNSSNAANTSTNSS 476

QY 593 ccctgctgccgcgctcgagcagagcctatgagacttctcggcgcgcatgcatcctacy 652
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 477 NTTTSSNNSSSCNTTSSANTTSTTTTATAGNNNNATSSNNNSNNNSNNNSNNSGSG 536

QY 653 gcacccgatacaaaaagAACCGCTCAAGCACGCCATCGAACCCTCTGGGCCCGCAGA 712
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Db 537 TSSGNTSSSCTGTGBSSGSNNSSNNNSNGSNNCTGTGSSNSSTTTTSSSSSSSAA 596
```

Search completed: October 6, 2002, 16:22:38  
Job time: 17806 sec







[illegible]

```

1  LENGTH: 6407 nucleotides
2  TYPE: nucleic acid
3  STRANDEDNESS: double
4  TOPOLOGY: circular
5  MOLECULE TYPE: nucleic acid
6  US-07-977-630-2

Query Match          2.3%  Score 82.2; DB 1: Length 6407;
Best Local Similarity 47.8%; Pred. No. 2,1e-10;
Matches 274; Conservative 0; Mismatches 293; Indels 6; Gaps 1;

QY 301 agagctggctgcgcacgcgaagccgcctgagcgttcacgcagagaagctctggggcgtacccggcagc 360
DB 3557 AGCTGTGCTGCGCCGACGACGACGCGCGCGCGGCAAGGACGACGATCTGCTCGAGGAGGATCTACCGCC 3616
QY 361 tgaccgcgcagagggcgcgtctgtagcctgccttataatgaaagcgaacccgltgtgtcagcagt 420
DB 3617 AAAGCCGGCGCGTGGCCCTTAAGGCCCGCCGGTACATGAGAGCGCAACCAACAGCGCTGGGCAA 3676
QY 421 ccttgatcatcacgcgatcgagatgtcttgatgtcgtacatgagccgcagacactcgtctgggc 480
DB 3677 ACCTCTGCTGGTGTGGACGATAGACATTCACAGACGACGCGTCCGACACGCTCAAGCGCCGGG 3736
QY 461 tgccttaccgctcactagctgtccatg-----aaccggtgtcaagcaacacggagacatcg 534
DB 3737 GGTCCATTCGCTGCGCCCAACGCGATCGTGGGCAATCGCGCCCAACGCGCACGACACGCGAG 3796
QY 535 tctatgcttgaagaagacactgtgtgtctgacccgatgcgcgcgcgcgcgcgcgcgcgcgcgcgc 594
DB 3797 TGTGGGCACTACCAACCCCTCTGTTCCACGACACCGAATACCGCGCGCTAAGCCGCTCGCAT 3856
QY 595 tgcctgcgcgcgtctcagagcaaggcctatgtcgacgttctcgcgcgcgcgcgcgcgcgcgcgcgc 654
DB 3857 ACATGGCGGCGCGCGCGCAAGAGCCTTTCGAGCGCGCGCTCGACGCGGACCGCACTGCTCAG 3916
QY 655 accgagatcacaaagaacccgcgtcagcagccgcgcacatgagcctctcggggccccgcgaagc 714
DB 3917 GCGTATATGACCAAAAACCCCGGCAATCGCTGGGAAAACGGAAAGGCTCTCCACTCAGATC 3976
QY 715 cgctttagagatgtgcgcgcctcgcacacacccctgcagagatctcagcaatctgcgcgaagc 774
DB 3977 TCTACACACTCAGCCACATCGAGGCGCGAGCTCGGGCGCAACTGTGCACCGCGCTCGGC 4036
QY 775 caagggaaccgcgcttcgcgaacgtcacccgaatcaagggctgcgcgcgaacgctacccctgtcg 834
DB 4037 GTCAACAACACACGATCAAAAGCGGCTCCGACGCCGCTATGGGGGGAATTGGCACTGTTCCG 4096
QY 835 aacacaccgcgatgtggtgcatccgggcgcgtcc 867
DB 4097 ATTCCGTCAGGTTGTGGGCTTATCTCCCGCCC 4129

RESULT 3
US-07-977-630-20
: Sequence 20, Application US/07977630
: Patent No. 5583038
: GENERAL INFORMATION:
: APPLICANT: Stover, Charles K.
: TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING
: TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS
: NUMBER OF SEQUENCES: 84
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Carella, Byrne, Bain, Giffillan, Cecchi,
: ADDRESSEE: Stewart & Olstein
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM

```

```

1 OPERATING SYSTEM: MS-DOS
2 SOFTWARE: ASCII
3
4 CURRENT APPLICATION DATA:
5 APPLICATION NUMBER: US/07/977, 630
6 FILING DATE: NO. 5583038ember 17, 1993
7
8 CLASSIFICATION: 435
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Herion, Charles J.
12 REGISTRATION NUMBER: 28, 019
13 REFERENCE/DOCKET NUMBER: 469201-174
14
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: 201-994-1700
17
18 TELEFAX: 201-994-1744
19
20 INFORMATION FOR SEQ ID NO: 20:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 4120 nucleotides
23 TYPE: nucleic acid
24 STRANDEDNESS: double
25 TOPOLOGY: circular
26
27 MOLECULE TYPE: nucleic acid
28
29 OS-07-977-630-20

```

Query Match	2.3%	Score 80.6	DB 1	Length 4120
Best Local Similarity	47.6%	Pred. No. 4.4e-10		
Matches 273	Conservative 0	Mismatches 294	Indels 6	Gaps 1

QY	301	agaagctgcgcgaacgaacccgtctgcctcaacgaagactctggcgctaccgacagc	360
Db	2382	AGCTCTGGCTGCGTACTGGCCGCTGGCAAGCAGCATGTGCTCCAGGGGATCTTACGGCC	244111
QY	361	tgactcgagcaagagcgctggaagctgtgctcttaacatgaaagcgaaaccgcttgctcatgcaagt	420
Db	2442	AAACGCGCGGCTGGCGCCCTAGCGCGCGCGGACATGAGGCAACCAACAGCGGCTGGCAA	2501
QY	421	ccctgtgcatcaacgcgatcgagatgtgcttgagatgtctgactggcgcgcaagactgcgttgac	480
Db	2502	ACCTGCTGGTGGTGGAGGTGAGCAATCCAGACGACGCTCCGACGCGTCAACGCGCCCGGG	2561
QY	481	tgacctcaacgcttactaagtgctcatg-----aacgctgtcaagcaacgagcaacatcg	534
Db	2562	GGTTCACATCCGCTGCGCCACGCGCATGTGGGCAATGCGCCCAACGGCCACAGCAGCGAG	2621
QY	535	tcctatgcttgaagaaaccctgtgtgtctacagatgctcgcgcgagcgagcgccctatcaacc	594
Db	2622	TGTGGGCACTTCACAGGCCCTGTCCACGCGACCGAATACGCGGGGTAAAGCCGCTCGCAT	2681
QY	595	tgctgcgcgcgtctcgagcaagcgctcatgcaagcttctcgcgcgagatgcatcctacagcgc	654
Db	2682	ACATGCGGGGTGCGCGCCGAGGAGGCTTGGCGCGCGCGTCGATGGGCACGCAAGTACTCAG	2741
QY	655	accggaatcaaaagaaccgcgtctagaacccgcacatgctgacacctctctggggcccgcaagcg	714
Db	2742	GCTTCATGACCAAAAACCCCGGCGCAACATGCGCTGGGAAACGGAAATGGCTCCACTCAATC	2801
QY	715	cgctctacgagctgcgcgcctctcgacaacacccctcgagagatccaagcaactgcccgaagc	774
Db	2802	TCTTACACACTCAGCCACATCGAAGCCGAGCTGGCGCGCAACATGCCACCGCGCGCTTGGC	2861
QY	775	caaggaaaccgcgctgcgaacgctcaaccgatacaacggttcgcccgaacgctcaaccctgttcg	834
Db	2862	GTCAGCGACACCAAGTACAAAGGCGGCTCCAGACGCGCTAGGGGGAATTGCGCATGTTCCG	2921
QY	835	acacacaccgcgaatgttggcgataccgagccgctgc	867
Db	2922	ATTCCGTCAGATTGTGGGCTATCGTCGCGGCC	2954

APPLICANT: Slover, Charles K  
 TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING  
 TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS  
 NUMBER OF SEQUENCES: 84  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Carella, Byrne, Balin, Giffillan, Cecchi,  
 ADDRESSEE: Stewart & Olstein  
 STREET: 6 Becker Farm Road  
 CITY: Roseland  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07068  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch diskette  
 COMPUTER: IBM  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/977,630  
 FILING DATE: No. 5583038emberr 17, 1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Herron, Charles J.  
 REGISTRATION NUMBER: 28,019  
 REFERENCE/DOCKET NUMBER: 4652201-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4120 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: nucleic acid  
 OS-07-977-630-21

Query Match	2.38;	Score 80.6;	DB 1;	Length 4120;
Best Local Similarity	47.68;	Pred. No. 4.4e-10;		
Matches 273; Conservative	0;	Mismatches 294;	Indels 6;	Gaps 1

[illegible]





```

1 GENERAL INFORMATION:
2 APPLICANT: Pathogenesis Corporation
3 APPLICANT: 201 Elliott Avenue West, Suite 150
4 APPLICANT: Seattle, Washington 98119
5 TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
6 TITLE OF INVENTION: Thereof
7 NUMBER OF SEQUENCES: 18
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Townsend and Townsend Khourie and Crew
10 STREET: Steuart Street Tower, One Market Plaza
11 CITY: San Francisco
12 STATE: California
13 COUNTRY: US
14 ZIP: 94105-1493
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: PCT/US95/12642
22 FILING DATE: 02-OCT-1995
23 CLASSIFICATION:
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 08/316,950
26 FILING DATE: 03-OCT-1994
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Hunter, Tom
29 REGISTRATION NUMBER: 38,498
30 REFERENCE/DOCKET NUMBER: 15377A-001100
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (415) 543-9600
33 TELEFAX: (415) 543-5043
34 INFORMATION FOR SEQ ID NO: 15:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 4296 base pairs
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (plasmid)
41 FEATURE:
42 NAME/KEY: misc.feature
43 LOCATION: 1..4296
44 OTHER INFORMATION: /standard_name= "plasmid pmh28"
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949

```

[illegible]

Accession	Sequence	Position
Db 2705	ACATGGGGGGTGGCGGAGAGGCTTGGCGGCGCGCTGCATGGCACCAGCACTTACTACG	2764
Oy 655	accggaatcaaaagaaccggtcagaacgcgccatgagacctctggyggcccgcaagc	714
Db 2765	GCCCTCATGACCCAAAACCCCGGCGACATGCGCTGGGGAAACGGAAATGGCTCCACTAGATC	2824
Oy 715	cgctctagagctgcygcccctcgcaacaacctcgagagatccagcactgcccggag	774
Db 2825	TCTACACATCAGCACATATGAGAGCCGAGCTGGCGCGAACAATGCAACCGCGCGCTGGC	2884
Oy 775	caaggaaaccgcytcgcgaacgltcaaccggaatcaacgyltcggccgcaacgltcaaccttlltcg	834
Db 2885	GTCAGCAGACACAGCTAACAAAGCGGCTCCAGCGCCGCTTAGGCGGGAATTGCGACTGTTCG	2944
Oy 835	acaccaccgcgatgtggcattaccggygcgcttc	867
Db 2945	ATTCCGTAGGTTGTGGGCTATCGTCCGCC	2977

```

      9
US-08-316-950-16
; Sequence 16, Application US/08316950
; Patent No. 5679515
; GENERAL INFORMATION:
; APPLICANT: STOVER, Charles K.
; APPLICANT: HICKEY, Mark J.
; TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Hourie and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,950
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15377A-001100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (plasmid)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..4352
; OTHER INFORMATION: /standard_name= "Plasmid pMH29"
US-08-316-950-16

Query Match          2.3%; Score 80.6; DB 1; Length 4352;
Best Local Similarity 47.6%; Pred. No. 4.5e-10;
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

    301 agagctgctcgcacacgaaccgcgttgccgtcagcccgagaagtctggggcgttacggcacg 360
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db   2405 AGCCTCGGCTGCCGCACGTCCGAGTGGCGCGGCAAGCAGCATCTGCTCGACGGAGATCTTCACGCC 2464
```

361 tgaatcgagagggcgctgagctgcttacaatgaagcgaacccgttgtaicagct 420  
2465 AAGCGCGCGCTGCGGCCCGCGGCTGATCATCGAGCGCAACCAACAGCGCTGCGCAA 2524  
421 ccttgatcaaccgatgagatgcttgatgctgagctgagcgccgagacccctgcctggcg 480  
2525 ACCTGCTGGTCTGTGAGCTAGACATCCATCGACGACGCTCCGAGCGCTCAGCGCGCGGG 2584  
481 tgcctcaacgctcactgctgctcagc-----aacggtgcaacgacccgagacatcg 534  
2585 GGTCCCATTCGGTGGCCCAACGGGATGTTGGCAATGCGCCCAAGCGCCACGACGACGAG 2644  
535 tctatgcttgaagaacctgtgtctgacgaatgacgagcgagcgagccatcaacc 594  
2645 TGTGGGCACTCAACCGCCCTGTTCCACGACCGCAATACCGCGCGCTAGCGCGCTGCAT 2704  
595 tctgagcccgctgagcgagcgacatgagatgctgagcttctgagcgagctgatacctaaggcg 654  
2705 ACATGGCGGCGTGGCGGAGGCTTGGGCGCGCGCTCATGCGCGCACTTACTCAG 2764  
655 accggaatcaagaacacccgctcagacacccgcatgcaaccccttgaggcccgacagcg 714  
2765 GCGTCAATGACAAAACCCCGGCGCATGCGCTGGAAACGGAATGCTCCTCAGATC 2824  
715 cgcttaagactgagcgccctcgaacacacccctcgaagagatccagacactgagcgagcg 774  
2825 TCTACACACTCAGCCATCGAGCGCGAGCTCGGCGCAATGCGCGCGCGCTGCG 2884  
775 cagggagacccgctgcaacgctcgaacccgatcaacgctgagcgagcgacgctgagcg 834  
2885 GTACAGACACACGTAACAAAGGCGCTCCGACGCGCTAGGGGGAATTCGCGCACTGTTGG 2944  
835 acacaccccgatgtaggacataccggcgctcc 867  
2945 ATTCCGTACAGTTGTGGCGCTATCGTCCGCC 2977

RESULT 10  
PCT-US95-12642-16  
Sequence 16, Application PC/7US9512642  
GENERAL INFORMATION:  
APPLICANT: Pathogenesis Corporation  
APPLICANT: 201 Elliott Avenue West, Suite 150  
APPLICANT: Seattle, Washington 98119  
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses  
TITLE OF INVENTION: The use of  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/12642  
FILING DATE: 02-OCT-1995  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,950  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE NUMBER: 15371A-001100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4352 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (plasmid)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..4352  
OTHER INFORMATION: /standard\_name= "plasmid pmh29"  
PCT-US95-12642-16  
Query Match 2.3%; Score 80.6; DB 5; Length 4352;  
Best Local Similarity 47.6%; Pred No. 4.5e-10;  
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;  
301 agagctgctgacacgacgacgacgctgagctgacgagagagcttgaggcgtaaccgacag 360  
2405 AGCTGTGGCTGCGCGGACGCGGACGCGCGCTGCGCAAGGACGATCGCTCGAGGGGATCTACCGCC 2464  
361 tgaatcgagagggcgctgagctgcttacaatgagcgaacccgttgtaicagct 420  
2465 AAGCGCGCGCTGCGGCCCGCGGCTGATCATCGAGCGCAACCAACAGCGCTGCGCAA 2524  
421 ccttgatcaaccgatgagatgcttgatgctgagctgagcgccgagacccctgcctggcg 480  
2525 ACCTGCTGGTCTGTGAGCTAGACATCCATCGACGACGCTCCGAGCGCTCAGCGCGCGGG 2584  
481 tgcctcaacgctcactgctgctcagc-----aacggtgcaacgacccgagacatcg 534  
2585 GGTCCCATTCGGTGGCCCAACGGGATGTTGGCAATGCGCCCAAGCGCCACGACGACGAG 2644  
535 tctatgcttgaagaacctgtgtctgacgaatgacgagcgagcgagccatcaacc 594  
2645 TGTGGGCACTCAACCGCCCTGTTCCACGACCGCAATACCGCGCGCTAGCGCGCTGCAT 2704  
595 tctgagcccgctgagcgagcgacatgagatgctgagcttctgagcgagctgatacctaaggcg 654  
2705 ACATGGCGGCGTGGCGGAGGCTTGGGCGCGCGCTCATGCGCGCACTTACTCAG 2764  
655 accggaatcaagaacacccgctcagacacccgcatgcaaccccttgaggcccgacagcg 714  
2765 GCGTCAATGACAAAACCCCGGCGCATGCGCTGGAAACGGAATGCTCCTCAGATC 2824  
715 cgcttaagactgagcgccctcgaacacacccctcgaagagatccagacactgagcgagcg 774  
2825 TCTACACACTCAGCCATCGAGCGCGAGCTCGGCGCAATGCGCGCGCGCTGCG 2884  
775 cagggagacccgctgcaacgctcgaacccgatcaacgctgagcgagcgacgctgagcg 834  
2885 GTACAGACACACGTAACAAAGGCGCTCCGACGCGCTAGGGGGAATTCGCGCACTGTTGG 2944  
835 acacaccccgatgtaggacataccggcgctcc 867  
2945 ATTCCGTACAGTTGTGGCGCTATCGTCCGCC 2977

RESULT 11  
US-08-316-950-13  
Sequence 13, Application US/08316950  
Patent No. 5679515  
GENERAL INFORMATION:  
APPLICANT: STOVER, Charles K.  
APPLICANT: HICKEY, Mark J.  
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses  
TITLE OF INVENTION: The use of  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza





Db 2562 GGTCCATCGCTGCCACGCGATGCGGCAATCGCGCAAGCGGCACACGCGAG 2621  
QY 535 tctatcgcttgaagaacccctgtgtctgacgatgcgagcgagcgacgtatcaacc 594  
Db 2622 TGTGGGCACTCAACAGCCCGCTGTCCACGACGCCAATACGGCGCGCGTAAGCCCTCGCAT 2681  
QY 595 tgcctgcgcgcgtcgacgagggcctatgacagcttcctcgagcgagatgcatcctacgggc 654  
Db 2682 ACATGCGCGGCTGCGCGGAGGCGCTTCGGCGCGCTCATGCGACCGCATGTTACTCAG 2741  
QY 655 accgagtcacaaagaacccgctcagacacgcccacatgcagacccctctgggcccgcagag 714  
Db 2742 GCGTCATGACAAAACCCCGGACATCGCTGGGAAAGGAATGGCTCCACTCAGATC 2801  
QY 715 cgcctacgagctgcgcgcctcgcacacacccctcgacagatcctcaacgacgtccgagag 774  
Db 2802 TCTACACACTCAGCCACATGAGGCGGAGCTCGCGCGGAACATGCCACCCCGCGCTGCG 2861  
QY 775 cagggaacccgcgtcgcaacgctacccgacatcaacggttcgagcgcaacgtcacctctg 834  
Db 2862 GTGAGGACAGCACGTAACAAAGCGGCTCCGACGCGGCTAGGGCGGAATTGCGCACTGTTCG 2921  
QY 835 acacacccgcagatgtggcataccggcgctcc 867  
Db 2922 ATTCCGTCAGGTGTGGGCTATGTCGCCGCC 2954

RESULT 13  
US-08-316-950-12  
Sequence 12, Application US/08316950  
Patent No. 5679515

GENERAL INFORMATION:  
APPLICANT: STOVER, Charles K.  
APPLICANT: HICKEY, Mark J.  
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316, 950  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 15371A-001100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 6047 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: DNA (plasmid)  
TOPOLOGY: linear

FEATURE:

NAME/KEY: misc.feature  
LOCATION: 1..6047  
OTHER INFORMATION: /standard\_name= "plasmid pMH30-lux"  
US-08-316-950-12

Query Match 2.3%; Score 80.6; DB 1; Length 6047;  
Best Local Similarity 47.6%; Pred. No. 5e-10;  
Matches 273; Conservative 0; Mismatches 294; Indels 6; Gaps 1;

QY 301 agagctgtgtcgcacgcaacgctgtgcacgcggaagatctgtggcgtacggcagc 360  
Db 2405 AGCTGTGCTGCTGCGGCTGCTGCGGCAAGCAGCATGCTGTCGAGGGGATCTACGCGC 2464  
QY 361 tgaactggcagagagcgctgtgagctgtcctaacatcgaaagaaacccgtgtgatgaat 420  
Db 2465 AAAGCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2524  
QY 421 ccttgatcacacacgagatgctgtcgtgatgtcgtgagcgagcagacccgtcgtggc 480  
Db 2525 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2584  
QY 481 tgcctacacgctcctcgtgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 534  
Db 2585 GGTCCATCGCTGCCACGCGATCGTGGGCAATCGCGCGCAACGCGCGCACACGCGAG 2644  
QY 535 tctatgcttgaagaacccctgt 594  
Db 2645 TGTGGGCACTCAACGCCCGCTGTCCACGACCGAATACGCGCGCGCTGATGCTGCTGCTG 2704  
QY 595 tgcctgcgcgcgtcgagcagcgacgtatgagcgttctcgcgcgcgcgcgcgcgcgcgcgcgc 654  
Db 2705 ACATGGCGGCTGCGCGGAGGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2764  
QY 655 accgagtcacaaagaacccgctcagacacgcccacatgcagacccctctgggcccgcagag 714  
Db 2765 GCGTCATGACAAAACCCCGGACATCGCTGGGAAAGGAATGGCTCCACTCAGATC 2824  
QY 715 cgcctacgagctgcgcgcctcgcacacccctcgacagatcctcaacgacgtccgagag 774  
Db 2825 TCTACACACTCAGCCACATGAGGCGGAGCTCGCGCGGAACATGCCACCCCGCGCTGCG 2884  
QY 775 cagggaacccgcgtcgcaacgctacccgacatcaacggttcgagcgcaacgtcacctctg 834  
Db 2885 GTGAGGACAGCACGTAACAAAGCGGCTCCGACGCGGCTAGGGCGGAATTGCGCACTGTTCG 2944  
QY 835 acacacccgcagatgtggcataccggcgctcc 867  
Db 2945 ATTCCGTCAGGTGTGGGCTATGTCGCCGCC 2977

RESULT 14  
PCT-US95-12642-12  
Sequence 12, Application PC/TUS9512642

GENERAL INFORMATION:

APPLICANT: Pathogenesis Corporation  
APPLICANT: 201 Elliott Avenue West, Suite 150  
TITLE OF INVENTION: Mycobacterial Reporter Strains and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/12642  
FILING DATE: 02-OCT-1995  
CLASSIFICATION:



```

Db 2742 GCCTCATGACCAAAACCCCGGCCACATCGCCTGGGAAACGGAATGGCTCCACTCAGATC 2801
OY 715 cgcctacgagctgagcgccctcgacacacccctcgacgagatccaagcactgcccgaag 774
Db 2802 TCTACACACTCAGCCACATCGAGGCGGAGCTGGGCGGACATGCCACCGCGCGCTGGC 2861
OY 775 cagggaaccgagctcgcaacgltcaaccgatcaacggtcgccgcaacgltcaaccctgttcg 834
Db 2862 GTCAGCAGACCACTACAAAGCGGCTCCGACGCGCTAGGGCGGAAATTGCGACTGTTCG 2921
OY 835 acaccaccgcatgtgggcatataccgggcccgtcc 867
Db 2922 ATCCGTCAGTTGTGGGCTATCGTCCCGCCC 2954

```

Search completed: October 6, 2002, 16:34:24  
 Job time: 17542 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:07 ; Search time 55.08 Seconds

(without alignments)  
611.027 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596  
Sequence: 1 MDSFETLFPESWLPKRKPLAS.....IAGSASRAHTRQOFLGLS 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1596	100.0	303	21	AAV44636
2	329	20.6	368	18	AAW14835
3	315	19.7	368	13	AAK20992
4	287.5	18.0	367	14	AAK34543
5	218	13.7	248	13	AAK29622
6	218	13.7	248	13	AAK37872
7	106.5	6.7	458	22	AAK35246
8	97.5	6.1	492	22	AAK35243
9	97	6.1	3232	22	AAK69873
10	96.5	6.0	492	22	AAK35245
11	95.5	6.0	492	22	AAK35244

12	89.5	5.6	610	22	AAK90741
13	89	5.6	354	22	AAE10136
14	89	5.6	838	22	ABB70548
15	88	5.5	905	22	ABG29396
16	87	5.5	350	22	AAU61038
17	87	5.5	706	22	AAU48553
18	86.5	5.4	392	22	AAU53468
19	86	5.4	489	22	AAK86250
20	85.5	5.4	473	22	ABG16462
21	85.5	5.4	1219	22	ABG12658
22	85.5	5.4	1279	22	AAW78969
23	85.5	5.4	1301	22	AAW79953
24	85.5	5.4	1907	22	ABG15511
25	85.5	5.4	1909	22	ABG15534
26	85.5	5.4	2013	22	ABG12657
27	84.5	5.3	1018	22	ABG63718
28	84.5	5.3	4292	22	ABG17060
29	84.5	5.3	4302	17	AAW00870
30	84.5	5.3	4302	19	AAW33396
31	84.5	5.3	4302	19	AAW23830
32	84.5	5.3	4302	21	AAV92719
33	84.5	5.3	4303	17	AAK90302
34	84.5	5.3	4303	21	AAV95558
35	84.5	5.3	4339	17	AAK87539
36	84	5.3	693	22	ABG25744
37	84	5.3	693	22	ABG28822
38	84	5.3	717	22	ABG71311
39	83.5	5.2	299	21	AAV91671
40	83.5	5.2	381	20	AAV60162
41	83.5	5.2	450	19	AAW80605
42	83.5	5.2	450	21	AAV94307
43	83.5	5.2	450	22	AAU37841
44	83.5	5.2	450	22	AAW01090
45	83.5	5.2	458	22	AAW25289

#### ALIGNMENTS

RESULT 1	
ID	AAV44636 standard; Protein: 303 AA.
AC	AAV44636;
XX	
DT	07-APR-2000 (first entry)
XX	
DE	Propionibacterium LMG 16545 protein-1.
XX	
KW	Propionibacterium LMG 16545; vector; plasmid; antigen; vaccine; enzyme; nutritional factor; growth factor; clotting factor; antimicrobial; drug; hormone; vitamin B12; animal feed; lactic acid bacteria; foodstuff; cheese.
OS	Propionibacterium freudenreichii LMG 16545.
XX	
PN	WO9967356-A2.
XX	
PD	29-DEC-1999.
XX	
PF	25-JUN-1999; 99WO-EP04416.
XX	
PR	25-JUN-1998; 98EP-0305033.
XX	
PA	(KONN) GIST-BROCADES BV.
XX	
PI	Pouwels PH, Van Lujk N, Jore JPM, Luiten RGM,
XX	
DR	WPI: 2000-136977/12.
XX	
PT	N-PSDB; AAK49691.
XX	
PT	Novel vectors containing Propionibacterium sequences, used to express homologous or heterologous proteins

*boos SEQ 2 have very function*

XX	Claim 16; Page 26; 54pp; English.
PS	
CC	The present amino acid sequence is encoded by the endogenous plasmid
CC	Propionibacterium LMG 16545 ORF-1, derived from Propionibacterium
CC	friederreichii deposited under CBS 101022 or CBS 101023. The vector has
CC	insertion sites for foreign DNA fragments and is capable of autonomous
CC	replication. It can be used safely as they remain extrachromosomal and
CC	are very stable. They can be used to express homologous or heterologous
CC	proteins, which may include antigens for use in vaccines, nutritional,
CC	growth and clotting factors, antimicrobials, vitamins (especially vitamin
CC	B12), enzymes, hormones and drugs. The Propionibacterium host cells are
CC	used in the production of animal feeds, manufacture of foodstuffs like,
CC	sauages and cheese, where they may be used instead of or in addition to
CC	lactic acid bacteria and in the manufacture of vitamin B12. The DNA
CC	sequence can also be used as a source of primers and probes.
SQ	
Sequence	303 AA:
Query Match	100.0%; Score 1596; DB 21; Length 303;
Best Local Similarity	100.0%; Pred. No. 1.5e-159;
Matches 303; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 MSSEFLPEEESLPRKPKPLASAEKSGVRYHTRGRALRPLETYENPVMOSLVITDRASD 60
DB	1 mstfelilfpeswlprrplasaeksgsyrvhtgralelpyieanplymgsylvitdrasd 60
OY	61 ADMADDLAQLPSPSYSMNRKVTTTGHIVYALKNPVCITDAARRRPINLARVEGGLDYVL 120
DB	61 adwaadlaqlpspsysmsnrkvtttghivylalknpvcitdaarrripinlarveggldcyl 120
OY	121 GGDASTGHRITKNPSTAHATTWGPADALVELRALHTDEIHALPEAGNPRRWNTSTV 180
DB	121 ggdaasyghrtknplstahtlwgpadaelyelralahltdeihalpeagpnrrwtstv 180
OY	181 GGNVTLFDRTMMAYAVNVHSMGCPRAEMHYTFEHTHLNETIINADEFAFGGLWEKL 240
DB	181 gnnvtlfddrtmmayavnvshswgpraeewhtvfenhlhlnetiaideftpglglnelk 240
OY	241 HLSRSISRWVRNFETPTFRAROKAISLRGASKGEGKGKGIASGASRRARTROOFLE 300
DB	241 hlsrsisrwvrnfetptfrarqkaalslrgaakgykggnhgiasgasrtahtrqfgle 300
OY	301 GLS 303     301 gls 303
DB	
RESULT 2	
AAM14835	
ID	AAW14835 standard; Protein: 368 AA.
NC	AAW14835;
DT	28-MAY-1997 (first entry)
DE	Protein encoded by M.rep ORF2 of plasmid pmW101.
KX	Plasmid pmW101; Mycobacteria; lipoprotein secretion signal sequence;
KW	antigen; bacterial vaccine; Lyme disease; Borrelia burgdorferi;
RW	surface protein.
XX	
OS	Synthetic.
PB	US5583038-A.
PD	10-DEC-1996.
PF	21-OCT-1991; 91US-0780261.
PR	17-NOV-1992; 92US-0977630.
PR	21-OCT-1991; 91US-0780261.

XX (MEDI-) MEDIMACNE INC.  
PA  
XX  
XX Stover CK;  
P1  
XX  
XX WPI: 1997-042315/04.  
DR N-PSDB: AAT64413.  
XX  
XX Mycobacteria expressing secretion signal of lipoprotein and  
PT heterologous antigen, esp. outer surface protein A or B of *Borrelia*  
PT burgdorferi - are used in the form of a live bacterial vaccines  
PT against Lyme disease  
XX  
PS  
XX Example 1: Column 41-44; 122p; English.  
XX  
XX The sequences given in AAM14834-36 are proteins which are encoded by  
CC the plasmid pmw101. This plasmid was used in the construction of a  
CC series of vectors used to produce the recombinant *Mycobacteri*a of  
CC the invention. The recombinant *mycobacteri*a of the invention are  
CC transformed with DNA encoding a polypeptide which comprises a  
CC lipoprotein secretion signal sequence and an antigen (Ag) heterologous  
CC to the *mycobacteri*a. The lipoprotein secretion signal causes the Ag  
CC to be produced as a lipoprotein. The *mycobacteri*a may be used in the  
CC form of a live bacterial vaccines against Lyme disease, where the  
CC bacteria express a surface protein of *Borrelia burgdorferi*, the  
CC causative agent of Lyme disease. This sequence is printed in the  
CC specification in the C-terminal to N-terminal orientation as  
CC determined from the corresponding DNA sequence.  
XX  
XX Sequence 368 AA;  
XX

Query Match	20.6%;	Score 329;	DB 18;	length 368;
Best Local Similarity	33.5%;	Pred. No. 6.8e-26;		
Matches 104;	Conservative 39;	Mismatches 131;	Indels 36;	Gaps 10;

```

OY      2 DSFEFLPEESLPRKPLASAE-KSGAVRHWTROALEPTLEAPLYMOSLVITDRASP    60
        ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      66 defeql-----wlpypiasdellleglytq-srasalgrtyeaarptanllvvdvahpd    120
OY      61 ADMADIALA-GLPSPSYVMNRVTTTGHIVYALKNPCLTDAAARRRPIINLLARVEQGLCD    118
        ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      121 aalelaeasgshpnaaygnrangnahnawalnappreiegartrkplymacaaegllrr    180
OY      119 VLGGDATGHRITNPNPSTAHATLMGPADALYTELRALAHTLDELTHALPEAGNPRRNYT--    176
        ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      181 avdgdrsyysgmktknpglhwetewlhds---lytlshleaegaumpprwrqttyx    236
OY      177 --RSVGNNVTLPFTTRMAARAVRHSMGVLAWE---HTVEPHILLMETIADDEFAT    231
        ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      237 aaiprlgrncalfdsvtlwatyrallmrylptrnvddlgfalyechnaeafrcndvcp    296
OY      232 GPLGINELKHLSRSISRW-----VMRN---FPPETPAROKAISLGASKGKEGSHKG    282
        ||| |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      297 gplpsdeerraiansiwrttkrsirwadglivyealisarsaalsrxga-----ar    348
OY      283 GASGASRRRA    292
        || |::|||
Db      349 taastvarra    358

RESULT      3
AAR20992
AAR20992 standard; Protein; 368 AA.
XX
AAR20992;
XX
03-JUN-1992 (first entry)
XX
Protein "e" encoded by mycobacterial plasmid pmw101.
XX
Polymerase chain reaction; mycobacterial promoter; kanamycin;
resistance; BCG; Bacille Calmette-Guerin; origin of replication.

```

```

XX OS Synthetic.
XX AAR34543
XX MO9201783-A.
XX PD 06-FEB-1992.
XX PF 09-JUL-1991; 91WO-US04833.
XX PR 16-JUL-1990; 90US-0553907.
XX PA (YESH ) EINSTEIN A COLLEGE.
XX PI (UTPI-) UNIV OF PITTSBURGH.
XX PI Jacobs WR, Hatfull G;
XX DR WPI: 1992-064943/08.
XX PF DNA site-specific integration into Mycobacteria - useful as
XX PT adjuvant in vaccines and as therapeutic agent for malaria,
XX PT Influenza, herpes and human immunodeficiency virus
XX PS Example 3; Fig 24; 82pp; English.
XX CC Plasmid pYUB12 was constructed by ligating a 5kb Sau3 fragment from
XX CC pAL5000 (contg. a replication origin from M.fortuitum) to BamHI-
XX CC digested pU666 (contg. an E.coli ori and kanamycin-neomycin
XX CC resistance sequences). A 2586bp HpaI-EcoRV fragment from pYUB12
XX CC (comprising the minimum sequence necessary for plasmid replication
XX CC in BCG) was ligated to PvuII-digested pYUB8 to form pYUB53. (Plasmid
XX CC pYUB8 is a pBR322 deriv. which includes an E.coli replicon and a
XX CC kanamycin resistance gene). Twelve restriction sites were removed
XX CC by digesting pYUB53 with AatI, EcoRV and PstI. To eliminate 792
XX CC bases of the (inactivated) tet gene, the plasmid was digested with
XX CC Nari and the 6407bp gel purified fragment was religated/
XX CC circularised. The resulting plasmid was designated pMV101.
XX CC The nucleotide sequence of plasmid pMV101 is printed in the
XX CC specification but the copy quality is too poor to allow the
XX CC sequence to be included on the GENESQ nucleotide database.
XX CC Three proteins are encoded by pMV101, i.e. the kanamycin resistance
XX CC protein and proteins "d" and "e" encoded by the complementary
XX CC strand. See also AAR20991 and AAR20993.
XX SQ Sequence 368 AA;

Query Match 19.7%; Score 315; DB 13; Length 368;
Best Local Similarity 33.2%; Pred. No. 2e-24; Mismatches 128; Indels 42; Gaps 12;
Matches 104; Conservative 39;

OY 2 DSFETLPESWLPKRRPLASAE-KSGAYRHVTRQRALELPYIEANPLVMOSLYITDSD 60
DB 66 defeql-----wlpypplasadllleglyrq-strasagrryileampalannllvddndp 120
OY 61 ADWADLA--GLPSPSYVMNRVTGTHIVALKNPVCLTDARRRPINLLARVEQGLCD 118
DB 121 aalraisaagsbplpnaivgnranghahavalnapyrteyarkkplaymaaceaglr 180
OY 119 VLGGSASVGHRTTKNPILSTAHNTIMGPADALVELRALATLDEIHNL-LEAGNPR----- 172
DB 181 avdgdrsygmltkwpgelawetewlhds---lytlshleaeigammp-----prwrgt 233
OY 173 -RNVRTSTVGRNVTLFDTRRMAYRAVRHSMGCPVAEME---HVEFHILNERTIAD 228
DB 234 tynaaprlpgrncalfdsrlywayrpalmrlylptnvnvlgfaylaecrranaeipnd 293
OY 229 FAYGPLGLNELKHLRSISRW-----VWRN---FTPTFRARQKRAISRGASGKGEGG 279
DB 294 vcpbplpdevalanslwrtilktsrliwadgjlvyaealisarqsaisrkga----- 346
OY 280 HKGSIASGASRRA 292
DB 347 -artaastavarta 358

```

```

RESULT 4
ID AAR34543
XX AAR34543 standard; Protein; 367 AA.
XX AC AAR34543;
XX DT 04-JUN-1993 (first entry)
XX DE Tet d gene product from pMV101 - a deletion mutant of pYUB125.
XX KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
XX KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
XX KW pertussis; malaria; influenza virus; CTL; herpes virus.
XX OS Mycobacterium.
XX FH Key Location/Qualifiers
XX FT Misc-difference 269
XX FT /note= "X undefined in specification"
XX FT Misc-difference 317
XX FT /note= "X undefined in specification"
XX PM WO9307897-A.
XX PM 29-APR-1993.
XX PD 21-OCT-1992; 92WO-US09075.
XX PR 21-OCT-1991; 91US-0780261.
XX PA (MEDI-) MEDIMUNE INC.
XX PI Stover CK;
XX DR WPI: 1993-152187/18.
XX DR N-PSDB; AAQ41316.
XX CC Expression vector for expressing protein or polypeptide in
XX CC mycobacterium - contg DNA sequences encoding lipoprotein
XX CC secretion signal and peptide heterologous to bacteria
XX CC fusion protein of lipoprotein heterologous to bacteria
XX PS Example 1; Fig 5; 86pp; English.
XX CC This sequence is the tet d gene product encoded by plasmid pMV101.
XX CC When the nucleotide was decoded, the sequence in the specification
XX CC was found to contain several deletions and insertions which upset
XX CC the reading frame. Also many codons were found to encode amino
XX CC acids contrary to the universal genetic code. In addition two amino
XX CC acids in the sequence in the specification were given as X and were not
XX CC further defined.
XX SQ Sequence 367 AA;

Query Match 18.0%; Score 287.5; DB 14; Length 367;
Best Local Similarity 33.3%; Pred. No. 1.6e-21;
Matches 98; Conservative 35; Mismatches 124; Indels 37; Gaps 11;

OY 2 DSFETLPESWLPKRRPLASAE-KSGAYRHVTRQRALELPYIEANPLVMOSLYITDSD 60
DB 66 defeql-----wlpypplasadllleglyrq-strasagrryileampalannllvddndp 120
OY 61 ADWADLA--GLPSPSYVMNRVTGTHIVALKNPVCLTDARRRPINLLARVEQGLCD 118
DB 121 aalraisaagsbplpnaivgnranghahavalnapyrteyarkkplaymaaceaglr 180
OY 119 VLGGSASVGHRTTKNPILSTAHNTIMGPADALVELRALATLDEIHNL-LEAGNPR----- 174
DB 181 v-dgdrsygmltkmpghlawetew-----lhedytlshleaeigamppprwrgt 231

```





XX OS Corynebacterium thermoaminogenes.  
XX PN EPI076094-A2.  
XX PD 14-FEB-2001.  
XX PF 11-AUG-2000; 2000EP-0117225.  
XX PR 12-AUG-1999; 99JP-0228391.  
XX PA (AJIN ) AJINOMOTO CO INC.  
XX PI Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y,  
XX PT Sugimoto S;  
XX DR WPI: 2001-193202/20.  
XX DR N-PSDB: AAF24442.  
XX PT New plasmids derived from Corynebacterium thermoaminogenes, useful for  
XX PT improving coryneform bacteria, which can grow at elevated temperatures,  
XX PT and for producing useful substances (e.g. L-amino acids)  
XX PS Claim 5; Page 20-21; 32pp; English.  
XX CC The present invention provides plasmids isolated from Corynebacterium  
XX CC thermoaminogenes encoding a rep protein. These are useful for improving  
XX CC Coryneform bacteria, which are capable of growing at elevated  
XX CC temperatures and of producing substances such as L-amino acids. The  
XX CC present sequence is the rep protein from a Coryneform bacterial plasmid.  
XX SQ Sequence 458 AA;

Query Match 6.7%; Score 106.5; DB 22; Length 458;  
Best Local Similarity 20.1%; Pred. NO. 0.026;  
Matches 60; Conservative 42; Mismatches 119; Indels 77; Gaps 14;  
QY 72 SPSYVSMNRVTTGGHIVALKNPVCLTDAARRRPNLRLARVEOGLCDVYLGDPASYGHRIT 131  
DB 137 apawlgvnpplsgkqqlwcl-dprfaegtsntrlllaateentmfgadqashls 195  
QY 132 KNPLESTAH---ATLM---GPADALYELRALAHTL-----DETHALPEAGNPRRNWTR 177  
DB 196 twplhvddpplayswhcqhnrvdldelmevartmtgskkprehagfssgrarieaar 255  
QY 178 STV--GRNVTLFDTTRMAYRA-----VRHSG--GPVAMERTVEHIT----- 217  
DB 256 kataeakalaaladatlptaleasgqlldgvrvlwaaegraar-detafhalvtvgqlka 314  
QY 218 --HLNETIADIDEF-----ATGPLG-----LNETKHLRSISRWVWNEFPTETFR 260  
DB 315 agerlkdkkildayeraynvagavagdgrepdlpmrdtqtnarrrayvakg----- 367  
QY 261 ARQKAISLRGSKGKEG-----GKRGIAISGASRRR-----HTRQOFLBGLS 303  
DB 368 --qptvsarsltqtsgrgrkalatmgrrgaatsnarrrwadpesdyarqtrlerlarams 423

RESULT 8  
AAB35243  
ID AAB35243 standard; Protein; 492 AA.  
XX AC AAB35243;  
XX DT 08-MAY-2001 (first entry)  
XX DE Corynebacterium thermoaminogenes plasmid pYU1 rep protein.  
XX KW Plasmid; pYU1; rep; coryneform bacterium; L-amino acid; thermostable;  
XX KW high temperature.  
XX OS Corynebacterium thermoaminogenes.

XX PN EPI076094-A2.  
XX PD 14-FEB-2001.  
XX PF 11-AUG-2000; 2000EP-0117225.  
XX PR 12-AUG-1999; 99JP-0228391.  
XX PA (AJIN ) AJINOMOTO CO INC.  
XX PI Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y,  
XX PT Sugimoto S;  
XX DR WPI: 2001-193202/20.  
XX DR N-PSDB: AAF24439.  
XX PT New plasmids derived from Corynebacterium thermoaminogenes, useful for  
XX PT improving coryneform bacteria, which can grow at elevated temperatures,  
XX PT and for producing useful substances (e.g. L-amino acids)  
XX PS Claim 1; Page 11-12; 32pp; English.  
XX CC The present invention provides plasmids isolated from Corynebacterium  
XX CC thermoaminogenes encoding a rep protein. These are useful for improving  
XX CC Coryneform bacteria, which are capable of growing at elevated  
XX CC temperatures and of producing substances such as L-amino acids. The  
XX CC present sequence is the rep protein from a Coryneform bacterial plasmid.  
XX SQ Sequence 492 AA;

Query Match 6.1%; Score 97.5; DB 22; Length 492;  
Best Local Similarity 20.2%; Pred. NO. 0.26;  
Matches 55; Conservative 42; Mismatches 102; Indels 73; Gaps 14;  
QY 72 SPSYVSMNRVTTGGHIVALKNPVCLTDAARRRPNLRLARVEOGLCDVYLGDPASYGHRIT 131  
DB 137 apawlgvnpplsgkqqlwcl-dprfaegtsntrlllaateentmfgadqashls 195  
QY 132 KNPLESTAH---ATLM---GPADALYELRALAHTL-----DETHALPEAGNPRRNWTR 177  
DB 196 twplhvddpplayswhcqhnrvdldelmevartmtgskkprehagfssgrarieaar 255  
QY 178 STV--GRNVTLFDTTRMAYRA-----VRHSG--GPVAMERTVEHIT----- 217  
DB 256 kataeakalaaladatlptaleasgqlldgvrvlwaaegraar-detafhalvtvgqlka 314  
QY 218 --HLNETIADIDEF-----ATGPLG-----LNETKHLRSISRWVWNEFPTETFR 260  
DB 315 agerlkdkkildayeraynvagavagdgrepdlpmrdtqtnarrrayvakg----- 367  
QY 261 ARQKAISLRG-----ASKG-----GKEGCHK 261  
DB 368 --qptvsarsltqtsgrgrkalatmgrrgrgqk 397

RESULT 9  
AAB69873  
ID AAB69873 standard; Protein; 3232 AA.  
XX AC AAB69873;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 36411.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.

```
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001MO-US09231.
PF
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX MPI: 2001-656860/75.
XX DR N-PSDB; ABL13976.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 36411; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 3232 AA:
SQ

Query Match 6.1%; Score 97; DB 22; Length 3232;
Best Local Similarity 22.3%; Pred. No. 4.7;
Matches 49; Conservative 35; Mismatches 70; Indels 66; Gaps 11;

OY 101 ARRRPIN--LLARVEOGL-----CDVLGDASYGHRITKNP-----LSTRAHATLWG 144
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 3009 akrrsgdglfsgqqlstreqldrvrsgslyggdahnghleespekkp1saaags19g 3068
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 145 PA-----DALYEL---BALAHTLDELTAH-LPEAGNPRRNVTSTVGRNVTLEDT 189
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 3069 dsckdvrrsvpdaiyetdsdeaaashqfgarsltprsinpq-vvast--nslprlpt 3125
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 190 TRMMAYRAVHNSWGGPVAEMEHVFEHIIHLNFTIADFERATGRLGNEKLHSRSISRW 249
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 3126 tgy-----gapltstprktsq-----salnhpsnlst- 3153
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

OY 250 VWRNFTPEFRAROKAISLRSKSGKSGHKGSIASGAS 289
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 3154 vsatgsaksaarytrspqlaaraaavsgagsgsgvgvgas 3193
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 10
AAB35245
ID AAB35245 standard; Protein; 492 AA.
XX
AC AAB35245;
XX
DT 08-MAY-2001 (first entry)
XX
DE Corynebacterium thermoaminogenes plasmid pYM3 rep protein.
XX
KW Plasmid; pYM3; rep; coryneform bacterium; L-amino acid; thermostable;
KW high temperature.
XX
OS Corynebacterium thermoaminogenes.
XX
PN EPI076094-A2.
PD
```

```
XX 14-FEB-2001.
PD
XX
XX 11-AUG-2000; 2000EP-0117225.
PF
XX
PR 12-AUG-1999; 99JP-0228391.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;
XX Sugimoto S;
PI
XX MPI: 2001-193202/20.
XX DR N-PSDB; AAF2441.
XX
XX New plasmids derived from Corynebacterium thermoaminogenes, useful for
PT improving coryneform bacteria, which can grow at elevated temperatures,
PT and for producing useful substances (e.g. L-amino acids) -
XX
PS Claim 4; Page 17-18; 32pp; English.
XX
XX The present invention provides plasmids isolated from Corynebacterium
CC thermoaminogenes encoding a rep protein. These are useful for improving
CC Coryneform bacteria, which are capable of growing at elevated
CC temperatures and of producing substances such as L-amino acids. The
CC present sequence is the rep protein from a Coryneform bacterial plasmid.
XX
XX Sequence 492 AA:
SQ

Query Match 6.0%; Score 96.5; DB 22; Length 492;
Best Local Similarity 21.0%; Pred. No. 0.33;
Matches 57; Conservative 41; Mismatches 101; Indels 73; Gaps 15;

OY 72 SPSYVSMNRVYTGHTHYALKNPVCULDARRRPINLLARVEGICDVGSDASYGHRIT 131
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 137 apawigvnpisgkqqlwci-dpvfaeglttsntrl1aattteemtrivgdaqafshrs 195
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 132 KNP1STAH--ATLW-----GPADALYELRALAHTL-----DEIHALPE--AGNPRRNVT 177
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 196 twplhvsddpctayswbhqnrvtldelmevartmtgskkprhahqefssgrar1ear 255
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 178 STV--GRNVTLEFTRRMAYRA-----VRHSWG--GPVAEMEHVFEHI----- 217
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 256 kateaakalaaladatlptaleasgdlldgyrvlwaagear-detafnalvtvyqlka 314
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 218 --HLNFTIADPE-----ATGPLG-----LNEKLHSRSISRWVWRNFTPEFR 260
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 315 agerlkdaklidayeraynvagavgadgrepdlpamcdrtqumarrvrayakg----- 367
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 261 AROKAISLRG---ASKG-----GKEGCHK 281
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 368 --qptvsarsgtqgsrgrkalatmtgrtggqk 397
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 11
AAB35244
ID AAB35244 standard; Protein; 492 AA.
XX
AC AAB35244;
XX
DT 08-MAY-2001 (first entry)
XX
DE Corynebacterium thermoaminogenes plasmid pYM2 rep protein.
XX
KW Plasmid; pYM2; rep; coryneform bacterium; L-amino acid; thermostable;
KW high temperature.
XX
OS Corynebacterium thermoaminogenes.
XX
PN EPI076094-A2.
XX
PD 14-FEB-2001.
PD
```

XX 11-AUG-2000; 2000EP-0117225.  
XX  
XX 12-AUG-1999; 99JP-0228391.  
XX  
XX (AJIN ) AJINOMOTO CO INC.  
XX  
XX Matsuzaki Y, Kimura E, Nakamatsu T, Kurahashi O, Kawahara Y;  
XX Sugimoto S;  
XX WPI: 2001-193202/20.  
XX DR N-PSDB; AA24440.  
XX  
XX New plasmids derived from Corynebacterium thermoaminogenes, useful for  
XX improving coryneform bacteria, which can grow at elevated temperatures,  
XX and for producing useful substances (e.g. L-amino acids) -  
XX  
XX Claim 4; Page 14-15; 32pp; English.  
XX  
XX The present invention provides plasmids isolated from Corynebacterium  
XX thermoaminogenes encoding a rep protein. These are useful for improving  
XX coryneform bacteria, which are capable of growing at elevated  
XX temperatures and of producing substances such as L-amino acids. The  
XX present sequence is the rep protein from a Coryneform bacterial plasmid.  
XX  
XX Sequence 492 AA;

Query Match 6.0%; Score 95.5; DB 22; Length 492;  
Best Local Similarity 21.2%; Pred. No. 0.42;  
Matches 58; Conservative 42; Mismatches 97; Indels 77; Gaps 16;

OY 72 SPSYSMNVTGGHIVYALKNPVCLTDARRRPINILARVEGGLCDVGLGDSYGRIT 131  
DB 137 apawgvnplsgkcoqlwcl-dpfaaegtsntclllaateemrrvvgadaafhris 195  
OY 132 KNP-----STAHATW-----GPADALVELRALAHTL-----DEIHALPE--AGNPRNV 175  
DB 196 rplhvfddptays--whcqhnrvidlidelmevarmtgskprkhahqefsgrrarlea 253  
OY 176 TSTV--GNVTLFDTTRMAYRA-----VRHSMG--GPAVEMHTVFEHT----- 217  
DB 254 atkataeaaaladclptaleasgdlldgyrvlwaegrear-detafrhalvgyql 312  
OY 218 ----HLNFTIADFE-----ATGPLG-----LNELKHSRSISRWWRNFTPER 258  
DB 313 kaagerlkakklidayerynvavagvaggdgrepldpmadrtqumarrvayakg----- 367  
OY 259 FRAROKAISLRG---ASKG-----GREGCHK 281  
DB 368 ----qptvaarsltqtsrgkakatamgrrgqk 397

RESULT 12

AAE90741  
ID AAG90741 standard; Protein; 610 AA.

XX AAG90741;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 4495.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.  
XX PR 07-APR-2000; 2000JP-0159162.  
XX PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX DR N-PSDB; AAH65960.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
XX mutation point of a gene, measuring expression of a gene, analysing  
XX expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 4495; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
XX are useful for identifying the mutation point of a gene derived from a  
XX mutant of coryneform bacterium, measuring expression amount and  
XX analysing the expression profile or expression pattern of a gene derived  
XX from Coryneform bacterium, and identifying a homologue of a gene derived  
XX from coryneform bacterium. Coryneform bacteria are useful for producing  
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX particularly L-lysine. The present sequence is a protein described  
XX in the exemplification of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from the  
XX European Patent Office.

XX Sequence 610 AA;

Query Match 5.6%; Score 89.5; DB 22; Length 610;  
Best Local Similarity 21.2%; Pred. No. 2.5;  
Matches 68; Conservative 37; Mismatches 115; Indels 101; Gaps 16;

OY 24 SGAYRHVYTORALEPLYEANPLVMS---LVITD--RQASDQMDADLAGLSPSYVSM 78  
DB 48 sytdengt-----pilygadkegyvlgldkynqjvedltgl-glydylf 93  
OY 79 NRTVTGHIY-----YALKNPV--CLTDARRRPINILARVEGGLCDVGLGD 123  
DB 94 tttlnhyavvqelfrgljydnngymikettlgalspstr---clpdrylegtcpcigtd 150  
OY 124 ASYGHRTKNPLSTFAHATLWGPADALY---ELRALAHTLDEIHALPEAGNPRARVTRST 179  
DB 151 gargdgcncngqldpadlmpvskingetpeltvehflidpalaee-----ltewl 204  
OY 180 VGR-----NVTLEFDTTRMAYRAVRH--SWG--GPAVEMHTVFEHILNFTII 225  
DB 205 kgredwrvnvlkfslnldldirp---tamsrdldwglpivpegvqdnakklywfdavv 261  
OY 226 ADEFATGPLGLNKLKLSISRWWRNFTPE-----TFRARQ 263  
DB 262 g-----ylsasi-ewayrsgdpeawrtfwmdpetksyfyngkdnltfhsq 306  
OY 264 KAISLRGASKGKEGCHKGI 284  
DB 307 wpaellgyagkgsgrgelgld 327

RESULT 13

AAE10136  
ID AAE10136 standard; Protein; 354 AA.

XX AAE10136;

XX 29-NOV-2001 (first entry)



OY 152 -----LRAHLAHTDEIHALPEAGNPRNRVT-----RSTVGARVTLFDTTRMA 194  
 Db 554 vqpgavttspdsvelprphcpqlvtrllsdenyrlsttgs-----pcthhh 608  
 OY 195 YRAVRHSMG-----PVAFMEHTVFEEHILNLTITADEFAT 231  
 Db 609 hhhphhaegtllqpleedeh-hhhhh-----hdeftt 640

## RESULT 15

ABG29396  
 ID ABG29396 standard; Protein: 905 AA.

AC ABG29396;

DE 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29387.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS93583.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID NO 59755; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensic, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 905 AA;

Query Match 5.5%; Score 86; DB 22; Length 905;  
 Best Local Similarity 24.7%; Pred. No. 6.4;  
 Matches 63; Conservative 27; Mismatches 87; Indels 78; Gaps 15;

OY 27 YRHVTRORALE-----LPYI-----EANPLVMOS--LYITDRRASDADMAADLAG 69  
 Db 171 frynrerllnkmcgllpyvapellkrrfhaepvdcgvlvtamlagelpw-----d 225  
 OY 70 LPSPS---YVSMNRVYTGHIYALKNPVCLDARRRPINLARVEQGLCDVLCGDASY 126  
 Db 226 qpsdscqgsdckekkt-----ylnpwkkidsa---pigtsqpn-----lsg--st 268  
 OY 127 GHRITKNPLSTAHATLWGPD--ALYELRALAHTLDEIHALPEAGNPRNRVTRSTVGRNV 184  
 Db 269 dhqtsvqqlpgeffslpnpsdaaalyetssgintlse-hgsehsghsktvahtsge-- 325  
 OY 165 TLFDTTRKMAIYAVRHSMGPRV---AEKEHTVFEEH-----IHLNETII----- 225  
 Db 326 -----haesehaasgepaatehaegeltvgqpsgeqpsgehlsaqeplselesge 375  
 OY 226 --ADEFATGPLGLANE 238  
 Db 376 qpsdcqpsgehsge 390

Search completed: October 4, 2002, 15:34:40  
 Job time: 93 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 28.21 Seconds

(without alignments)  
1032.083 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596

Sequence: 1 MDSEFTLPESWLPKPLAS.....IAGSARRAHTRQGLEGLS 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	329	20.6	336	2 JS0052	37K protein - Myco
2	282.5	17.7	297	2 S04455	hypothetical prote
3	282.5	17.7	304	2 S04456	hypothetical prote
4	198.5	12.4	248	2 S32701	hypothetical prote
5	100	6.3	99	2 S32700	hypothetical prote
6	98	6.1	1461	2 B70588	probable polyketid
7	97.5	6.1	1876	2 E97944	zinc metalloprotei
8	93.5	5.9	228	2 H72703	probable translati
9	89.5	5.6	606	2 T37217	probable ATP/GTP b
10	89	5.6	1216	2 T09224	spindle assembly c
11	89	5.6	4558	2 C82199	RTX toxin RtxA Vc1
12	88.5	5.5	303	2 T36509	probable molybdopt
13	87.5	5.5	409	2 T14611	alanine transamina
14	87	5.5	457	2 S03321	regulatory protein
15	87	5.5	487	2 S60675	hypothetical prote
16	86.5	5.4	497	2 T14609	alanine transamina
17	86	5.4	672	2 JC4637	transketolase (EC
18	86	5.4	1217	2 T03457	transketolase (EC
19	85.5	5.4	1366	2 T00270	hypothetical prote
20	85.5	5.4	1366	2 T35985	probable large pro
21	85	5.3	356	2 H98304	limb2-dependent al
22	85	5.3	356	2 AE2378	alkanesulfonate mo
23	85	5.3	409	2 T06767	cellulase (EC 3.2.
24	85	5.3	900	2 JH0157	mtag protein (Afl8
25	84.5	5.3	355	2 H98275	monooxygenase (imp
26	84.5	5.3	355	2 AE3008	probable integral
27	84.5	5.3	500	2 T36090	polycystic kidney
28	84.5	5.3	4302	2 A38971	proteinase II (imp
29	84	5.3	686	2 C90948	

30	84	5.3	686	2 G85796	proteinase II (imp
31	84	5.3	1693	2 AC3240	helicase, SNF2 fam
32	83.5	5.2	450	2 E95257	replicative DNA he
33	83.5	5.2	450	2 E98122	replicative DNA he
34	83.5	5.2	863	1 GMYH12	retrovirus-related
35	83.5	5.2	1815	2 S73021	polyketide synthas
36	83.5	5.2	1822	2 F87203	polyketide synthas
37	83	5.2	251	2 I57999	SP-10 - western ba
38	83	5.2	285	2 I77964	probable acetolact
39	83	5.2	559	2 G83024	exodeoxyribonuclea
40	83	5.2	639	2 D82809	oligopeptidase B (
41	83	5.2	686	1 E64946	retrovirus-related
42	82.5	5.2	814	1 GNM51P	cobn protein homol
43	82.5	5.2	1116	2 D97695	cobalamin biosynth
44	82.5	5.2	1116	2 AC2921	hypothetical prote
45	82	5.1	497	2 T47715	

## ALIGNMENTS

RESULT 1  
JS0052  
37K protein - Mycobacterium fortuitum plasmid PAL5000  
C:Species: Mycobacterium fortuitum  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 10-Dec-1999  
C:Accession: JS0052; J01440  
R:Rauber, J.; Moniz-Pereira, J.; Gloquel-Sanzey, B.  
Gene 71, 315-321, 1988  
A:Title: Complete nucleotide sequence of PAL5000, a plasmid from Mycobacterium fortui  
A:Reference number: JS0052; MID:89138007  
A:Accession: JS0052  
A:Molecule type: DNA  
A:Residues: 1-336 <RAD>  
A>Note: the authors translated the initiation codon GTG for residue 1 as Val  
A>Note: the authors translated the codon ACA for residue 18 as Trp  
R:Labidi, A.; Mardis, E.; Roe, B.A.; Wallace Jr., R.J.  
Plasmid 27, 130-140, 1992  
A:Title: Cloning and DNA sequence of the Mycobacterium fortuitum var fortuitum plasm  
A:Reference number: J01440; MID:92311251  
A:Accession: J01440  
A:Molecule type: DNA  
A:Residues: 160-336 <LAB>  
A:Cross-references: GB:M60875; NID:g149984; PIDN:AAA25372.1; PID:g149985  
A:Experimental source: var. fortuitum  
C:Genetics:  
A:Genome: plasmid  
A:Start codon: GTG  
C:Function:  
A:Description: this protein is involved in the initiation of plasmid replication  
C:Superfamily: Mycobacterium hypothetical 37k protein

Query Match 20.6%; Score 329; DB 2; Length 336;

Best local similarity 33.5%; Pred. No. 5e-20;  
Matches 104; Conservative 39; Mismatches 131; Indels 36; Gaps 10;

QY	2	DSEFTLPESWLPKPLASAE	KSGATRYTORALELPTEANPLWQSLVTTDRDASD	60
DB	34	DEFEQTL---WLPWPLASDDLLEGIYRQ-SRASALGRRYEENPFLAMLVVDVDDHP	88	
QY	61	ADMAADLA--GLPSPSYVSNRVTTGHIYVALKNPCLDAAARRPINALVAREGLCD	118	
DB	89	AALRALASRSHPLPNAIVGNRANGAHAYALNAPVPRTEYARRKFLAYMAACAGGLRR	148	
QY	119	VLCGDSYGRITKPNPLSTAHAATLWGPADALYELRALAHTLDEIHALPEAGNPRNVT--	176	
DB	149	AVYDGRISYSLMTKPNPGHIMETEMLHSD---LYTISHIEALEGAMNPPRRRQOTYK	204	
QY	177	--KSTYGRNVTLDPTTMAAYRAVRHSGCPVAEME--HTEVBHILHMETTIADFEAT	231	
DB	205	AAPPLIGRNCALPDSVLMAYRPALMRLIYLPTRNVDELGRAIYAECCHARNAEPFCNDVCP	264	







```

Db      91  --HROSQVNAVLPVDSAGIAVSRSAVAVEG--ADAAGLGDCLCRKLTGTLPLAASAAALG 147

QY      144  GPAAALTELRLATLTDEIHALPLAGCPKRNVRITRYGKRVLTLEDITRMAYRAVRHSG 203
          | : : : : :
Db      148  VEAVAAVEESYKI-----VSNAQWEGV---ETR--SYEV--EGWG 180

QY      204  GPVAEMETVEEHILNETIADFGATGPIGLNEIK 240
          | : | | | : | | | : | : |
Db      181  ----DVEYEPVPEPIPLAVSTIATELOGGACGAVPEPLR 213

```

```

RESULT      9
T37217
Probable ATP/GMP binding protein - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37217
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21615
A:Accession: T37217
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-606 <OLIT>
A:Cross-references: EMBL:AL031514; PIDN:GMA20595.1; GSPDB:GN00070; SCOEDB:SC2H4.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC2H4.01

```

Query Match	5.6%	Score 89.5	DB 2	Length 606
Best Local Similarity	22.3%	Pred. No. 14		
Matches	67	Conservative	38	Mismatches 127; Indels 69; Gaps 16
QY	30	VTRORALELPYEANPLVMOSLVITDRDASDADMAADLAGLSPSPVSMNRVTYTGHIY	89	
Db	347	ITQOQLLEEGVPANELVAQ-----WDAASTGTYKELSSYAPSL-----LSAGAVLD	398	
QY	90	ALKNPVCLTDA-----RRRPNLLARVEQGLCDVIGDASGCHRTKNPLSTA	138	
Db	399	EENNALVYDGSAGYASATGPPVDESGFTVSARVOLDSAKLASKPVGVEALVAGQOASAG	458	
QY	139	HAT--LWG--PADALYE---LRALNHTIDETIALDEAGNPRNRYRSTYGRVYT--LEDTT	190	
Db	459	ESSMALNAKVPADGVYQKFTTFVAGADGKAYQASAE--PGGDIAEFTDITWVOTGVFDQA	516	
QY	191	RMNAVRAVRHSGGPN--AEWENYVEFHIHL--INFTIINDEPATGPIGLNELKHLRSIS	247	
Db	517	EPW-----EMTDPADSAKTE--TLYGGLHLHYGVGEFDQPSFVAG-----	553	
QY	248	RMWVNRNTEPTTFRRAROKAISLRGASKGKEGGHKG-----IASGASHRAHTROOFLEG	301	
Db	554	-----FTTPQYGAGALAVG--RG--SRGGTGGHHLRGALAEELRVMTGAMGADQVRQVLAG	605	
QY	302	L	302	
Db	606	V	606	

```

RESULT 10
T09224
spindle assembly checkpoint protein SLDA - Emeritella nidulans
C:Species: Emeritella nidulans, Aspergillus nidulans
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09224
R:Efimov, V.P.; Morris, N.R.
Submitted to the EMBL Data Library, November 1997
A:Reference number: Z16619
A:Accession: T09224
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1216 <FT>

```

A:Cross-references: EMBL:AF0332987; NID:g31336093; PDB:g31336094  
C:Experimental source: strain R153  
C:Genetics:  
A:Gene: slgA  
A:Introns: 331/3; 477/1  
C:Keywords: microtubule; mitosis

	Query Match	5.6%;	Score 89;	DB 2;	Length 1216;	
	Best Local Similarity	21.5%;	Pred. No. 37;			
	Matches	66;	Conservative	93;	Indels 106;	Gaps 15;
QY	17	PLASAEKGAAYRHVTRORALDELPLYEANPLVMQSLVTD-----	ROASDADMDAD	66		
DB	736	PLANKSGSSSLR-----VPSPTKDSNP--QGFIKQICLNCPIDRSIRODTITQDELHTT		785		
QY	67	LAGLPS-----ESYVSMNRVTTTGHLYVALKNVCLTDAAARRPFINLLARVEGGLCDVLGG	122			
DB	786	LAAPGYAHNP-----TQSH--YAPEIERFAKSSSKRSRSGGEAFVPIIDPPGG	835			
QY	123	DASYGHRITKNDLST-AHATLWGPADALYELRALNHTLDEIHALPEAGCPRRNRYTRSPG	181			
DB	836	ERST--IIRRELGCAYAPVY-----LAEISDINDSDSEMEVGSNSGRSTVS	881			
QY	182	RNVLTLEPTRMWYRAVYVSHSWGPAWEHTEVFEIH-----	218			
DB	882	NSLTFQKRP-RYSFAILKLEVGPRNA-WEFYVIGQTAHNRHLSQPLTSLRAADSIYRAHEMH	939			
QY	219	LNLETTIADEF-----ATGPIGLNELKHLNRSISRWVRNFTP	256			
DB	940	IFKNESILVEYRPOGTLLDVLNLRNEGISGSPATGEGGLDE-----SLAWF-----FTI	989			
QY	257	ETFRARQ	263			
DB	990	ELFRTIQ	996			

```

RESULT 11
C82199
RTX toxin RtxA VC1451 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82199
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gattin, M.L.; Dodson, R.;
charadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82053; MUID:20406833
A:Accession: C82199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4558 <HEI>
A:Cross-references: GB:AE004223; GB:AE003852; NID:9655942; PIDN:AAF94608.1; GSPDB:GNP
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1451
A:Map position: 1

```

Query Match	5.6%	Score 89	DB 2	Length 4558
Best Local Similarity	23.6%	Pred. No. 2e+02		
Matches	50	Conservative	20	Mismatches 76; Indels 66; Gaps 9
QY	124	ASVGRIRKNP-----LSTAHATLWGPADALY-----ELRALAHPLDETH	163	
Db	2386	ASKGPPPKYKNSKSSASAHKKAEMLTQAHQDAKTAFTYDNNHETLTSTKLANSKKEIH	2445	
QY	164	ALPEAGNDRRVNTRSTVGR--NVTLFDTPTRMAYRAVHR-----SMGSPVAEMEHVFE	215	
Db	2446	ALGDSNLENKKQGNVAASIINTMLNDMLPFALTEERNLVQBEDDEGFVPAWAGCTEDK	2505	
QY	216	HIHLINETIIADFEATGDLGELNKLHRSISRWWRNF-----TPETFRAROKAISLR	269	

Db 2506 -----STILED-----PEDAQKAIERFILANFDNEQMPDELFVNDKVIS-- 2550

QY 270 GASKGKEGKHKGASGASRAHTRROOFLG 301

Db 2551 -----HHEG-----RTHVLAQKVDG 2565

RESULT 12

T36509

probable molybdopterlin-guanine dinucleotide blosynthesis protein - Streptomyces coelicol

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T36509

R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: 221608

A:Accession: T36509

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-303 <SAU>

A:Cross-references: EMBL:AL096822; PIDN:CAB46951.1; GSPDB:GN00070; SCQEDB:SCGD3.29C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCQEDB:SCGD3.29C

Query Match

Best Local Similarity 23.5%; Score 88.5; DB 2; Length 303;

Matches 75; Conservative 33; Mismatches 104; Indels 107; Gaps 18;

QY 2 DSEETLPESWLPKPKPLAS---AEKSGAYRHVTRORAL---ELPYTEANPLV---MOSL 51

Db 44 DPRTAPYVWARDPGGGGLALAGLRHTTAEVYVASADLPFL-AGPTVGRLLSAL 102

QY 52 VTIDRDA---SDADWADLAGLPSPSYVSMNRVTTHGHYVALNPPCLTDAARRPINTL 108

Db 103 AAGGADSVLLTDAD-----GRDP-----LVAAYRAP-----ALRRELA 137

QY 109 LARVEGLCDVLGDDASYGHRITKNPLSTAHATLMGPADALYELRALHTLDELHALEP 168

Db 138 LARRRAG-----KGERADGAOLTELP-----LRRLTGAL-RLTRVPDA 174

QY 169 GNPARNVTRSTGVGRNVTLFTTRMAYRAVHSGGVPAAEMHTVFEHILNETIAD 228

Db 175 -----VAAPFC-----DTW-DLATARRARIREGHVLEWISA-- 206

QY 229 PANGPLGLN-----ELKHLRSRISRWVWRNFTP-ETFRAROKAISLGASKGEGCHK 281

Db 207 -AKDELGIDIDVDVTGILLDLARDAHGVARPAAPLTFVLGYAA-----GRAGGP 256

QY 282 GGASGASRRHTRROOFL 300

Db 257 EAVAEETARKAALAAQNAE 275

RESULT 13

T14611

alanine transaminase homolog - Trypanosoma cruzi

C:Species: Trypanosoma cruzi

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Jun-2000

C:Accession: T14611

R:Andersson, B.; Aslund, L.; Pettersson, U.

submitted to the EMBL Data Library, March 1998

A:Description: 93.4 kb of complete sequence from chromosome 3 of Trypanosoma cruzi.

A:Reference number: 218159

A:Accession: T14611

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-409 <AND>

A:Cross-references: EMBL:AF052832; NID:g3063540; PID:g3063548; PIDN:AACT14084.1

C:Genetics:

A:Map position: 3

C:Superfamily: alanine transaminase

Query Match

Best Local Similarity 23.3%; Score 87.5; DB 2; Length 409;

Matches 80; Conservative 36; Mismatches 127; Indels 101; Gaps 16;

QY 15 RKPLASAE-----KSGAYRHVTRORAL-----LPYTEANPLVQSLVITDRD 57

Db 33 RAALATPEGKAKYPFSSIVYCNIGNPQALEGKPLTFEFRQVSLDA-PFLEENKRYTSQ 91

QY 58 ASDADMAA-----DLACLPSPSYSVSMNRVTTHGHYVAL-KNPPCLTDA 100

Db 92 PADAVARREYLRHIGDTGAYTSAGYAFARDIVARQINRDEHRIPLVYASSIFLTDG 151

QY 101 ARRPINLVAEGLCDVLGDDASYGHR---ITKNPLSTAHATLMGPADALYEL----- 152

Db 152 A-SSGVRLILO-----VLVGDSADAVMVPPIPOPLTAQTLTGTPAMYIYCEKDNW 203

QY 153 ---RALAHTLDEI---HALPEA-----GNPRNVTRSTGVGRNVTLFTTRMAYRAV 198

Db 204 ALNVEELASVYDECYAKNNATPRVLVYINPGNPGVLDROVMEVAKFCDDR----- 256

QY 199 RNSGCPYAEWEHTVFEHILNETIADPERATGPLGELNEKHLRSISRVMWRNFTPET 258

Db 257 -----GIVLMADEVYQENVYAGKRFLS--FREYVLGL-----PAP 290

QY 259 FRAROKAISLGASKG---GKEGKHKGASGASRAHTRROOFL 300

Db 291 YNTDTVLASLHSTSKGIIG-ECGRRGYFCLTNPPAPVREYVK 333

RESULT 14

S03321

regulatory protein ntrC - Rhodobacter capsulatus

N:Alternate names: regulatory protein nifH1

C:Species: Rhodobacter capsulatus

C>Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 02-Feb-2001

C:Accession: S03321; S37299; S34983

R:Jones, R.; Haselkorn, R.

MOL. Gen. Genet. 215, 507-516, 1989

A:Title: The DNA sequence of the Rhodobacter capsulatus ntrA, ntrB and ntrC gene anal

A:Accession number: S03321; M01D:89218961

A:Reference number: S03321

A:Molecule type: DNA

A:Residues: 1-457 <JON>

A:Cross-references: EMBL:X12359; NID:g46077; PIDN:CAA30922.1; PID:g46079

R:Kranz, R.

submitted to the EMBL Data Library, March 1993

A:Reference number: S37299

A:Accession: S37299

A:Molecule type: DNA

A:Residues: 1-457 <KRA>

A:Cross-references: EMBL:X72382; NID:g313162; PIDN:CA51075.1; PID:g313166

R:Poster-Hartnett, D.; Cullen, P.J.; Gabbert, K.K.; Kranz, R.G.

MOL. Microbiol. 8, 903-914, 1993

A:Title: Sequence, genetic, and lacZ fusion analyses of a nifH3-ntrB-ntrC operon in R

A:Reference number: S34980; M01D:93360820

A:Accession: S34983

A:Molecule type: DNA

A:Residues: 334-449 <FOS>

A:Cross-references: EMBL:X72382

C:Genetics:

A:Gene: nifH1

C:Superfamily: nitrogen assimilation regulatory protein ntrC; response regulator homo

C:Keywords: DNA binding; nucleotide binding; P-loop; phosphoprotein; transcription re

F:5-114/Domain: response regulator homology <RRA>

F:143-336/Domain: RNA polymerase sigma factor interaction domain homology <SFI>

F:171-178/Region: nucleotide-binding motif A (P-loop)

F:53/Binding site: phosphate (asp) (covalent) #status predicted

Query Match

5.5%; Score 87; DB 2; Length 457;

Best Local Similarity 24.6%; Pred. No. 15; Matches 75; Conservative 28; Mismatches 100; Indels 102; Gaps 16;

Search completed: October 4, 2002, 15:35:15  
Job time: 127 sec

OY 5 ETLFESWLPKPLA--SAKSGAYRHVROALELPI-----EANPLVMSLVTDR 56  
DB 132 ETSRPEG--TDLPVGRRTAMQALRYVARVMADLPVIMGESGKSLIAKAT----H 185  
OY 57 DASD-----ADMADLAGLPSPSYVMNV-----TTTGHTIV---ALK 92  
DB 186 DFSDRRTLPFVVAQAADLLGADGPSLLARQGRVVEFEDVDYDDETQGRIVRMIDLAP 245  
OY 93 NPV-----CLTQARRRRPINLLARVQGLCDVIGSDASYGHRITKNPPLSTAHAT 141  
DB 246 DPPRIMATTQVDLGLAMEGR-----RFRDLYYRLGG-----VT 280  
OY 142 LMGPADALYELRALAHTLDEIHALPE--AGNPRNVTRSTVGRNVTLPDTRMMAYRAVR 199  
DB 281 LAVP-----ALRERVEDIPLAEHFLGRAERD-----GLGMRAFSAEAMGLVRA-- 324  
OY 200 HSMGCPVAEMEHTEHHLNET-----IIADEPATGPLGL-NELKHLRSISRW 249  
DB 325 YAMPGNVROLENTVRLVVTASEEITRTVEFVLGNCPAVEPLGAGGEEKLSASIAH 384  
OY 250 VWRNF 254  
DB 385 LRRYP 389

RESULT 15

S60675  
hypothetical protein C - Corynebacterium glutamicum plasmid pGAL  
C:Species: Corynebacterium glutamicum  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 22-Oct-1999  
C:Accession: S60675  
R:Neavera, J.; Palek, M.; Hochmannova, J.; Abrahmova, Z.  
submitted to the EMBL data library August 1995  
A:Description: Complete nucleotide sequence of the cryptic plasmid pGAL from Corynebacte  
A:Reference number: S60673  
A:Accession: S60675  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-487 <NES>  
A:Cross-references: EMBL:X90817; NID:g951006; PIDN:CAA62330.1; PID:g951009  
C:Genetics:  
A:Genome: plasmid pGAL

Query Match 5.5%; Score 87; DB 2; Length 487;

Best Local Similarity 20.2%; Pred. NO. 17; Mismatches 116; Indels 72; Gaps 13;

OY 72 SPSYVSMNRVTTHIYVYALKNPVCLTDAARRRPINLLARVQGLCDVIGSDASYGHRT 131  
DB 138 APAMIGVNPPLSGKQOLW-LIDPYVAAAGKTSNNMRLLAATTEEMTRVFGADQAFSHRLS 196  
OY 132 KNPULSTAH---ATLW---GPDALYELRALAHTLDEIHALPEAGNPRNRV----- 175  
DB 197 RMPLEHVSDDPTAYKWHCOHDRVRLADLMEIART-----MTGSQPKPKYIODESSGRA 250  
OY 176 -----TRSTV-GRNVTLPDTRFMAYRA-----VRHSMGCPVAEMEHTEHTEH----- 217  
DB 251 RLEAQRATAEKALATLDASLPSALDSGLIDGVRVLMWNPERRADETAFRHALTVGY 310  
OY 218 -----HLLNETIIDEF-----ATCGPLG-----LNELKHLRSISRWVNRNFTP 256  
DB 311 QLKAAGERLKDAAKITDAYEVAVNVAQVAGADGREDPLAMRDLTMARVRGVYAKG-QP 369  
OY 257 ETPRAROKAISLRGASGKGKGGKGTASGASRA-----HTROQPLE 300  
DB 370 VVPARRVETOSRG-RKALATMGRRGAATSNARRWADEPSKYAQTRORLAE 420



**THIS PAGE BLANK (USPTO)**



CC -1- TISSUE SPECIFICITY: ISOFORM SREBP-1C PREDOMINATES IN LIVER.  
 CC ADRENAL GLAND, BRAIN AND ADIPOSE TISSUE, WHEREAS ISOFORM SREBP-1A  
 CC PREDOMINATES IN SPLEEN, BOTH ISOFORMS ARE FOUND IN KIDNEY, THYMUS,  
 CC TESTIS, MUSCLE, JEJUNUM, AND ILEUM.  
 CC -1- PIM: Under sterol-depleted conditions, SREBPs are cleaved  
 CC sequentially by site-1 and site-2 protease. The first cleavage  
 CC occurs within the luminal loop and is directly regulated by  
 CC sterols. The second cleavage by site-2 protease occurs within the  
 CC first transmembrane stretch and liberates the soluble  
 CC transcription factor. Cleavage by the cysteine proteases, caspase-  
 CC 3 and caspase-7, is induced during apoptosis, independent of  
 CC sterol levels.  
 CC -1- MISCELLANEOUS: SREBPs have to be in a complex with the cleavage-  
 CC activating protein (SCAP) to move to the Golgi and be cleaved by  
 CC site-1 protease.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB017337; BAA74795.1; -  
 DR EMBL; A1060765; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; A1132428; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AA475250; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AA498625; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; A1386259; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; A1552467; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; A1595781; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; A1786503; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AF374266; AAK54762.1; -  
 DR HSSP; P36956; IAm9.  
 DR MGD; MG1:107606; Strebf1.  
 DR InterPro; IPR003015; HLH\_Myc.  
 DR InterPro; IPR001092; HLH\_dlm.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SMO0353; HLH; 1.  
 DR PROSITE; PS00038; Helix\_LOOP\_HELIX; 1.  
 KW Transcription regulation; Activator; DNA-binding; Lipid metabolism;  
 KW Cholesterol metabolism; Nuclear protein; Transmembrane;  
 KW Endoplasmic reticulum; Golgi stack; Alternative splicing.  
 FT DOMAIN 1 477  
 FT TRANSMEM 478 498  
 FT DOMAIN 499 536  
 FT TRANSEM 537 557  
 FT DOMAIN 558 1134  
 FT DOMAIN 1 60  
 FT DVA\_BIND 317 330  
 FT DOMAIN 331 368  
 FT DOMAIN 367 388  
 FT DOMAIN 420 453  
 FT SITE 451 452  
 FT SITE 480 481  
 FT SITE 519 520  
 FT VARSPIC 1 29  
 FT VARSPIC 90 131  
 FT CONFLICT 272 276  
 FT CONFLICT 510 510  
 FT CONFLICT 613 613  
 FT CONFLICT 621 621  
 FT CONFLICT 625 625  
 FT CONFLICT 633 633  
 FT CONFLICT 906 906  
 FT CONFLICT 906 906

SO SEQUENCE 1134 AA; 120505 MW; E0DF184FF2A4FB7D0 CRC64;  
 Query Match 5.6%; Score 89.5; DB 1; Length 1134;  
 Best Local Similarity 21.3%; Pred. No. 6.3;  
 Matches 74; Conservative 38; Mismatches 120; Indels 115; Gaps 17;  
 OY 5 ETLPESEWLPKPKPLASAKSGAYRHVTRORALELPYIEANPLWQSLYTRDASDA--- 61  
 DB 774 ESLYSVACNPVYDLAQVYR--LFRHLLERLALNC-IAQSPQADG---DREFSDALGY 826  
 OY 62 ---DMADLAGLPSPY-VSNMRYTTG----- 85  
 DB 827 LQLNCSGDAGAPACSFSSVSSMATTPPVAKWMSLAAYVIMLRDEEAERLYP 886  
 OY 86 ---HI-----YALKNPCLND--AARRPIVL-----LARV 112  
 DB 887 LVEHLPQYLDTERPLPRAHLXSFKAARALDHRVESPASALCEKASGLRSLAST 946  
 OY 113 EOG-----LCDVL--GGDASYGHRITKNPLSTAHTLGMGADALYELRALAHTLD 160  
 DB 947 PTGSSIDRAMOLLCDLLVARTSLMQQSPASVQVAHGTSNGPOASALELRFQHDLS 1006  
 OY 161 EIHALPEAGNP-----RANVTSTYGRNLTLEDTRMAYRAVRISMGPVAE- 208  
 DB 1007 SLRLAQSFPRMRRVFLHEATRLMAGASPARTHQLD--RSLRRAGSSGKGTTAEL 1064  
 OY 209 -----W-EHTVEFHILNFTIIDEFATGP-IGLNELKHLSTSR 248  
 DB 1065 EPRPWREHT---EALLLASCLLPAPFASBQGRMSMLAEARTYK 1108  
 RESULT 2  
 NTRC\_RHOCA STANDARD; PRT; 457 AA.  
 AC P09432;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Nitrogen regulation protein ntrc.  
 GN NTRC OR NTRF1.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum;  
 CC Rhodospirillum.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SB1003;  
 RX MEDLINE-89218961; PubMed-2710108;  
 RA Jones R., Haselkorn R.;  
 RT "The DNA sequence of the Rhodospirillum rubrum ntrc and ntrc  
 RL Mol. Gen. Genet. 215:507-516(1989).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SB1003;  
 RX MEDLINE-93360820; PubMed-8355615;  
 RA Foster-Hartnett D., Cullem P.J., Gabbett K.K., Kranz R.G.;  
 RT "Sequence, genetic, and lacZ fusion analyses of a ntrc3-ntrb-ntrc  
 RL operon in Rhodospirillum rubrum."  
 MO1. Microbiol. 8:903-914(1993).  
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM NTRB/NTRC  
 CC INVOLVED IN THE ACTIVATION OF NITROGEN ASSIMILATORY GENES SUCH AS  
 CC GUNA. NTRC IS PHOSPHORYLATED BY NTRB AND INTERACTS WITH SIGMA-54.  
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.  
 CC INTERACTION ATP-BINDING DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X12359; CAA30922.1; -  
 CC DR EMBL; X72382; CAA51075.1; -  
 DR PIR; S03321; S03321.  
 DR HSP; P11028; LEV.  
 DR InterPro; IPR003593; AAA.  
 DR InterPro; IPR002197; HTH\_Fls.  
 DR InterPro; IPR001789; Response\_reg.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR Pfam; PF02854; HTH\_8; 1.  
 DR Pfam; PF00072; response\_reg; 1.  
 DR Pfam; PF00158; Sigma54\_activat; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR SMART; SM00448; REC; 1.  
 DR PROSITE; PS00110; RESPONSE\_REGULATORY; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; FALSE\_NEG.  
 DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
 DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
 DR Nitrogen fixation; Transcription regulation; Repressor; Activator;  
 KW DNA-binding; ATP-binding; Phosphorylation; Sensory transduction.  
 FT DOMAIN 1 118  
 FT MOD\_RES 53 53 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 171 178  
 FT NP\_BIND 143 343 SIGMA-54 FACTOR INTERACTION (POTENTIAL).  
 FT DNA\_BIND 425 444 H-T-H MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 457 AA; 50042 MW; 7EB04BAFE4740306 CRC64;

Query Match 5.5%; Score 87; DB 1; Length 457;  
 Best Local Similarity 24.6%; Pred. No. 3.3;  
 Matches 75; Conservative 28; Mismatches 100; Indels 102; Gaps 16;

QY 5 ETLPPSWLPKRPPLA--SAEKSQAYHNVTQRALSLPYT-----EANPLVMOSLYITDR 56  
 DB 132 ETSRPG--TDLPVLRGTAMQALYRLVARVMNADLPVIMGESGTSKSLAKAI----H 185  
 QY 57 DMSD-----ADMADLGLDPSVSNMRY-----TTTGHIVY---ALK 92  
 DB 186 DSDRTLPFVYQAADLDGADGPFSSLLARQGRVLVFEVDGDDYDETQGRIVRMADALP 245  
 QY 93 NPV-----CLTDARRRRPINLLARVEGLDVLGDSASGHRITTKNPLSTAHAT 141  
 DB 246 DPRIWATQVDLGLALMEGR-----REFODLYYRLGG-----VT 280  
 QY 142 LMGPADALYELALAHITLDEIHAPL--AGNPRRNVTRSTGVGNVTLPFTTRMAYRAVR 199  
 DB 281 LAVP-----ALREVEDIPLAEHFLGAEERD-----GIGMRAFSANAGLYRA-- 324  
 QY 200 HSMGPFVAEMHTVFEHILNLT-----IIDEFATGPIGL--NELKHLRSISIRP 249  
 DB 325 YAMPGVNROLENTVRLVLTATSEETTRTEVEFVLGNQPAVEPLGAGGEKLSIAIRH 384  
 QY 250 VWRNF 254  
 DB 385 LRRYF 389  
 RESULT 3  
 TKT\_RHOCA STANDARD; PRT; 672 AA.  
 AC 052723; 068024;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Transketolase (EC 2.2.1.1) (TK).  
 GN TKTA.  
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 CC Rhodobacter.  
 NCBI\_TaxID=1061;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33303 / B10;  
 RX MEDLINE-96186907; PubMed-8635754;  
 RA Sully D'Aspremont R., Toussaint B., Vignais P.M.;  
 RT Isolation of Rhodobacter capsulatus transketolase: cloning and  
 RL sequencing of its structural tkla gene.";  
 RN gene 169:81-84(1996).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SB1003;  
 RX MEDLINE-97404404; PubMed-9256491;  
 RA Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Feinstein M.;  
 RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter  
 RT capsulatus SB1003";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 CC -1- CARBOLYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-gluceraldehyde  
 CC 3-phosphate = D-ribose 5-phosphate + D-xylulose 5-phosphate.  
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.  
 CC -1- PATHWAY: KEY ENZYME OF THE REDUCTIVE AND THE OXIDATIVE PENTOSE  
 CC PHOSPHATE PATHWAYS.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L48803; AAB06805.1; -  
 DR EMBL; AF010496; AAC16110.1; -  
 DR HSP; P23254; TTR.  
 DR InterPro; IPR000360; Transketolase.  
 DR Pfam; PF00456; transketolase; 1.  
 DR Pfam; PF02780; transketolase; C; 1.  
 DR Pfam; PF02779; transketolase; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE; 2; 1.  
 KW Transferrase; Thiamine pyrophosphate.  
 FT CONFLICT 376 382 PESTIGS -> LNRHRL (IN REF. 1).  
 FT CONFLICT 484 484 A -> T (IN REF. 1).  
 SQ SEQUENCE 672 AA; 72436 MW; B6B1E7DE1B66CA80 CRC64;

Query Match 5.4%; Score 86; DB 1; Length 672;  
 Best Local Similarity 22.2%; Pred. No. 6.6;  
 Matches 61; Conservative 37; Mismatches 103; Indels 74; Gaps 14;

QY 63 WAADLGLDPSVSNMRY--TTTGHIVYALKNVCLDARRRPINLLARVEGLD--- 118  
 DB 103 WGARMMG--HPEYGHLEGEVETTTGPGGISTAVGMAIAK-----SMAAFGRKLVDHKT 156  
 QY 119 -VLGDASYGHRITKNPLSTAHATLWGPADALYELALAHITLDEIHAPLPEAGNPRRNVTR 177  
 DB 157 WVLADGDCMEGTSQEAIGLAG-----KQEDNLVIMDNNN----- 193  
 QY 178 STYGRNVTLFDTTRMAYRAVRHSMGQVAEME-----HTVEHILN----- 221  
 DB 194 ITIDGRVTQSDVTDOKARFA-----ASGMVDLSCDGHADIDIRALTAAKARPYL 245  
 QY 222 ---ETII-----ADEPAT--GPIGLNELKHLRSISIRVVMNFT--PEFFRAKOAISL 268  
 DB 246 VDCKTILIGGSPKADSIYAHGAPLGDALTK-LTRRAYGHEHGFVYIPATIKAEWALGA 304  
 QY 269 RGASKGKGEHNGKIASGASRRRAHTRQOFLKLS 303  
 DB 305 KGAABRAEMEARLALPAG--KRAEFERQMARQVA 337

RESULT 4  
 GDNH\_C10TM

ID	GENE	STANDARD	PRT	900 AA
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase)			
DE	(Cellulase H).			
GN	Cellulase H.			
OS	Clostridium thermocellum.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1515;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NCIB 10682;			
RK	MEDLINE=90323606; PubMed=2197182;			
RA	Yaguee E., Beguin P., Aubert J.-P.;			
RT	"Nucleotide sequence and deletion analysis of the cellulase-encoding			
RT	gene celH of Clostridium thermocellum.";			
RL	Gene 89:61-67(1990)			
CC	-1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-			
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-			
CC	GLUCANS.			
CC	-1- CATALYTIC ACTIVITY: Endohydrolaysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS			
CC	WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN			
CC	MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.			
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY			
CC	A (FAMILY 5 OF GLYCOSYL HYDROLASES).			
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 26 OF			
CC	GLYCOSYL HYDROLASES.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-slb.ch/announce/slb.ch">http://www.isb-slb.ch/announce/slb.ch</a> ).			
CC	or send an email to <a href="mailto:license@slb-slb.ch">license@slb-slb.ch</a> ).			
CC	-----			
DR	EMBL: M31903; AAA23325.1; -			
DR	PIR: JH0157; JH0157.			
DR	HSSP: P07985; ICED.			
DR	InterPro: IPR002105; Dockerin_1.			
DR	InterPro: IPR002048; EF-hand.			
DR	InterPro: IPR001547; Glyco_hydro_F5.			
DR	Pfam: PF00150; cellulase; 1.			
DR	Pfam: PF00404; Dockerin_1; 2.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.			
DR	PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.			
FT	SIGNAL	1	44	
FT	CHAIN	45	900	ENDOGLUCANASE H.
FT	DOMAIN	45	630	CATALYTIC (BY SIMILARITY).
FT	DOMAIN	631	904	PRO/THR-RICH (LINER).
FT	DOMAIN	655	900	CELLULOSE-BINDING (BY SIMILARITY).
FT	ACT_SITE	460	460	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	565	565	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	833	895	2 X 24 AA APPROXIMATE REPEATS.
FT	REPEAT	833	856	
FT	REPEAT	872	895	1.
FT	REPEAT	872	895	2.
SEQUENCE	900 AA; 102415 MW; 973AFB1954FC246B CRC64;			
Query Match	5.3%;	Score 85;	DB 1;	Length 900;
Best Local Similarity	22.5%;	Pred. NO. 12;		
Matches	48;	Conservative 26;	Mismatches 75;	Indels 64;
				Gaps 12;
113	EGGLCDVAGGA-----SYGRIIRKNPLSTAHATITMGADALYELR-----			153
95	EYNTVDINKNGADAVITRMAQDMKRAYGEIWRPLRHEANGD-WYPMWAGYSSRVNTNEY			153

Oy		154	--ALAHITDEIHALPEACNPR--RWVTSYGRNVTLF-----DTPMAAYR-AVR 196     :      :
Dd		154	IAAFEHIYD-IFRANGANNKWFERNVDNGNSTSYLGHPGDNYYDYTSDIXMGTT 212       :     :
Oy		200	HSMGGPVAWE-----HVFEEHILLNETLIADFPATGPLGLNELKHLSRSISR----248     :     :
Dd		213	OSWG---SOMOSFDQVFRAOYLASINKPIIIIEFAAEIGNKAKWITEAYSIRTSY 265 ::: ::  :     :
Oy		249	-----VV-----WR-NFTPEFRAROKAI 266   :      :     :
Dd		270	NKVIAAVWFHENKETMDWRINSSPALAAYREAL 302     :     :
RESULT	5		
P_KDL_HUMAN	STANDARD;	PRT:	4303 AA.
ID_P_KDL_HUMAN			P98161;
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Polycystin precursor (Autosomal dominant polycystic kidney disease protein I).		
Gn	PKDI.		
Os	Homo sapiens (Human).		
Cc	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Cx	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
Rn	[1]		
Rp	SEQUENCE FROM N.A.		
Rx	MEDLINE=95254638; PubMed=7736581;		
Ra	Glaesemann-Kuis M.A., Tayber O., Woolf E.A., Bougueleret L.,		
Ra	Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,		
Ra	Duyk G., Schneider M.C., Feng L., Zhang F., Zhao Z., Torosian S.,		
Ra	Readers S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,		
Rt	Nowicka U., Leung A.L.S., Frischaut A.-M.;		
Rt	"Polycystic kidney disease: the complete structure of the PKD1 gene		
Rl	and its protein.";		
Rl	Cell 81:289-298(1995).		
Rn	[2]		
Rp	SEQUENCE OF 2769-4303 FROM N.A.		
Rx	MEDLINE=94273192; PubMed=8004675;		
Ra	Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,		
Ra	MacCarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,		
Ra	Higgs D.R., Ratcliffe P.J., Harris P.C., Roelstema J.H.,		
Ra	Spruit L.L., Saris J.F., Dauwere H.G., Peters D.J.M.,		
Ra	Bruning M.H., Nellist W., Brook-Carter P.T., Maheshwar M.M.,		
Ra	Cordeiro T., Santos L.W., Cabral P., Sampson J.R., Janussen B.,		
Ra	Hesselink-Janssen A.L.W., van den Omweland A.M.W., Bussen B.,		
Ra	Verhoef S., Lindhout D., Hailey D.J.J.;		
Rt	"The polycystic kidney disease 1 gene encodes a 14 kb transcript and		
Rt	lies within a duplicated region on chromosome 16. The European		
Rt	Polycystic Kidney Disease Consortium."		
Rl	Cell 77:881-894(1994).		
Rn	[3]		
Rp	STRUCTURE BY NMR OF 275-354.		
Rx	MEDLINE=99107746; Pubmed=9889186;		
Ra	Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,		
Ra	Thomas R.L., Choctha C., Clare J., Hamill S.J., Sandford R.,		
Rt	"The structure of a PKD domain from polycystin-1: Implications for		
Rt	polycystic kidney disease."		
Rl	EMBO J. 18:297-305(1999).		
Rn	[4]		
Rp	VARIANT ADPPK 3748-ARG-VAl-3752 DEL, AND VARIANT ASP-3632.		
Rx	MEDLINE=96108969; Pubmed=8554072;		
Ra	Petal B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,		
Ra	Strong C., Harris P.C.;		
Rt	"Screening the 3' region of the polycystic kidney disease 1 (PKD1)		
Rt	gene reveals six novel mutations.";		
Rl	Am. J. Hum. Genet. 58:86-96(1996).		
Rn	[5]		
Rp	VARIANTS ADPPK P-2993; R-3016 AND V-3511, AND VARIANTS M-3510 AND		
Rp	F-4190.		

RA MEDLINE-97342914; PubMed-9199561;  
RA Peral B., Gamble V., Strong C., Ong A.C.M., Sloane-Stanley J.,  
RA Zervas K., Winearis C.G., Harris P.C.;  
RT "Identification of mutations in the duplicated region of the  
RT polycystic kidney disease 1 gene (PKD1) by a novel approach.";  
RL Am. J. Hum. Genet. 60:1399-1410(1997).  
RN [6]  
RP VARIANT ALA-4058.  
RX MEDLINE-97295081; PubMed-9150733;  
RA Constantinides R., Xenophontos S.L., Neophytou P., Nomura S.,  
RA Pierides A., Constantinou-Deltas C.D.;  
RT "New amino acid polymorphism, Ala/Val4058, in exon 45 of the  
RT polycystic kidney disease 1 gene: evolution of alleles.";  
RL Hum. Genet. 93:644-647(1997).  
RN [7]  
RP VARIANTS T-2760; P-2761; V-2763; T-2764; Q-2791; T-2826; L-3008 AND  
RP L-3064.  
RX MEDLINE-97449169; PubMed-9285784;  
RA Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandalph M.A.,  
RA Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;  
RT "An unusual pattern of mutation in the duplicated portion of PKD1 is  
RT revealed by use of a novel strategy for mutation detection.";  
RL Hum. Mol. Genet. 6:1473-1481(1997).  
RN [8]  
RP VARIANT ADPKD THR-3678.  
RX MEDLINE-97403939; PubMed-9259200;  
RA Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,  
RA Pignatti P.F.;  
RT "Three novel mutations of the PKD1 gene in Italian families with  
RT autosomal dominant polycystic kidney disease.";  
RL Hum. Mutat. 10:164-167(1997).  
RN [9]  
RP VARIANT ADPKD ASP-4032 AND VARIANT VAL-4045.  
RX MEDLINE-98180892; PubMed-9521593;  
RA Danielli C., Maheshwar M.M., Lazarou L., Davies F., Coles G.,  
RA Ravine D., and recurrent mutations in the PKD1 (polycystic kidney  
RT disease) gene.";  
RL Hum. Genet. 102:216-220(1998).  
RN [10]  
RP VARIANT ADPKD MET-3375.  
RX MEDLINE-99118881; PubMed-9921908;  
RA Koptides M., Constantinides R., Kyriakides G., Hadjigavriel M.,  
RA Patalis P.C., Pierides A., Deltas C.C.;  
RT "Loss of heterozygosity in polycystic kidney disease with a missense  
RT mutation in the repeated region of PKD1.";  
RL Hum. Genet. 103:709-717(1998).  
RN [11]  
RP VARIANTS ADPKD L-324 AND S-845, AND VARIANTS R-1399 AND L-1786.  
RX MEDLINE-99294580; PubMed-10364515;  
RA Thomas R.L., McConnell R., Whitlacker J., Kirkpatrick P., Bradley J.,  
RA Sandford R.;  
RT "Identification of mutations in the repeated part of the autosomal  
RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range  
RT PCR.";  
RL Am. J. Hum. Genet. 65:39-49(1999).  
RN [12]  
RP VARIANTS ADPKD P-2392 AND F-2423, AND VARIANTS R-1399; Q-2548 AND  
RP R-2638.  
RX PubMed-10577909;  
RA Watnick T., Phakdeekitcharoen B., Johnson A., Gandalph M., Wang M.,  
RA Briel G., Klinger R.W., Kimberling W., Gabow P., Germino G.G.;  
RT "Mutation detection of PKD1 identifying a novel mutation common to  
RT three families with aneurysms and/or very-early-onset disease.";  
RL Am. J. Hum. Genet. 65:1561-1571(1999).  
RN [13]  
RP VARIANTS ADPKD L-3994--F-3996 DUP, G-4136 AND C-4154, AND VARIANTS.  
RX MEDLINE-20441957; PubMed-10987650;  
RA Perichot R.A., Mercier B., Simon P.M., Whebe B., Clodes J., Ferec C.;  
RT "DGE screening of PKD1 gene reveals novel mutations in a large cohort  
RT of 146 unrelated patients";  
RL Hum. Genet. 105:231-239(1999).  
RN [14]  
RP VARIANTS ADPKD 3748-R-V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.  
RX MEDLINE-20112427; PubMed-10647901;  
RA Aizal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,  
RA Jeffery S.;  
RT "Novel mutations in the 3' region of the polycystic kidney disease 1  
RT (PKD1) gene.";  
RL Hum. Genet. 105:648-653(1999).  
RN [15]  
RP VARIANTS ADPKD PRO-4225 AND TRP-4276.  
RX MEDLINE-99217041; PubMed-10200984;  
RA Badenas G., Torra R., San Millan J.L., Lucero L., Milla M.,  
RA Estivill X., Darnell A.;  
RT "Mutation analysis within the 3' region of the PKD1 gene.";  
RL Kidney Int. 55:1225-1233(1999).  
RN [16]  
RP VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS CYS-2379; LEU-3066;  
RP VAL-3139 AND LEU-3193.  
RX MEDLINE-20311156; PubMed-10854095;  
RA Perichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,  
RA Clodes J., Ferec C.;  
RT "Novel mutations in the duplicated region of PKD1 gene.";  
RL Eur. J. Hum. Genet. 8:353-359(2000).  
RN [17]  
RP VARIANTS ADPKD PRO-2921 AND MET-3375, AND VARIANT LEU-3066.  
RX MEDLINE-20382887; PubMed-10923040;  
RA Koptides M., Mean R., Demetriou K., Constantinides R., Pierides A.,  
RA Harris P.C., Deltas C.C.;  
RT "Screening of the PKD1 duplicated region reveals multiple single  
RT nucleotide polymorphisms and a de novo mutation in Hellenic  
RT polycystic kidney disease families.";  
RL Hum. Mutat. 16:176-176(2000).  
RN [18]  
RP VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.  
RX MEDLINE-20514565; PubMed-11058904;  
RA Aguiari G., Savelli S., Garbo M., Bozza A., Angello G., Penolazzi L.,  
RA De Paoli Vitale E., La Torre C., Cappelli G., Piva R., del Senno L.;  
RT "Novel splicing and missense mutations in autosomal dominant  
RT polycystic kidney disease 1 (PKD1) gene: expression of mutated  
RT genes.";  
RL Hum. Mutat. 16:444-445(2000).  
RN [19]  
RP VARIANTS ADPKD SER-1166; GLU-1956; CYS-2408 AND GLY-2442--2443 INS,  
RP AND VARIANTS HIS-1995 AND ASN-2604.  
RX MEDLINE-20467506; PubMed-11012875;  
RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,  
RA Germino G.G.;  
RT "Thirteen novel mutations of the replicated region of PKD1 in an Asian  
RT population.";  
RL Kidney Int. 58:1400-1412(2000).  
RN [20]  
RP VARIANTS ADPKD TRP-3753 AND ASN-3815.  
RX MEDLINE-20275386; PubMed-10729710;  
RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,  
RA Hwang D.Y., Lee J.G., Namkoong Y., Lee C.C.;  
RT "Novel mutations of the PKD1 gene in Korean patients with autosomal  
RT dominant polycystic kidney disease.";  
RL Mutat. Res. 432:39-45(2000).  
RN [21]  
RP VARIANTS ADPKD Q-13; F-75; C-139; 1992-F-1993 DELINS L; 2220-R--P-  
RP 2224 DEL; D-2336; D-2752; 2762-TMR-2765 DUP; M-2768; K-2771; P-2816;  
RP S-2858; 3012-T--Y-3011 DEL AND 3748-L--R-3752 DEL, AND VARIANTS S-  
RP 2674; M-2708; T-2734; L-2735; C-2765; M-2782; R-2814; G-2888; I-2905;  
RP D-2966 AND L-3066.  
RX MEDLINE-21063179; PubMed-11115377;  
RA Rossetti S., Scimecki L., Gamble V., Burton S., Sneddon V., Peral B.,  
RA Roy S., Bakraloglu A., Komel R., Winearis C.G., Harris P.C.;  
RT "Mutation analysis of the entire PKD1 gene: genetic and diagnostic  
RT implications";  
RL Am. J. Hum. Genet. 68:46-63(2001).  
RN [22]  
Query Match 5.3%; Score 84.5; DB 1; Length 4303;  
Best Local Similarity 21.1%; Pred. No. 95;

Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps 9;

OY 4 FETTESWLPKPKPLASAEKSGAYRHV-----TRQALPELYEANPLVQSLVI 53  
DB 1391 FVOLDDEAML-----VACAMPPEPRYRTWDFGTGEAAPTRARQPEVTFEYRDE-----GSYLV 1443

OY 54 TDRQASDADMDADLAGLSPSPVSYNMRYTTHGHLYALKNPCLDAAARRRINLARVE 113  
DB 1444 TVTASNNISAAANDSALVEQEPVLTSLKVGSLGLEQOPLTFVAGVGRPASV----- 1498

OY 114 QGLCDVLGDASVGHRTKPLNSTAHATL-----WGPADALYELRALAHTLD 160  
DB 1499 --LMDLGGCMHEGPEVYHAINSTQDFYVAGMNEVSRSEMLVYTKRRVGLV----- 1552

OY 161 EIHALPEAGNPRRYRTSTVGNHTLFTPTRMAYRAVRHSN-----GGPVAEME 210  
DB 1553 -----VNASRTVPLNGSVSFTSLGSDVRYSWLDCRCPPIGCPITISV- 1599

OY 211 HTVEFHILNLTITADEFATGCPGLNFKLHSRISRMV 250  
DB 1600 --TFRSVGTFTNIVTAE-----NEVGSADSIIFYV 1628

## RESULT 6

POL\_IPHA STANDARD; PRT; 863 AA.

AC P04026;  
DT 23-OCT-1986 (Rel. 02, Created)  
DT 23-OCT-1986 (Rel. 02, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative POL polypeptide [Contains: Endonuclease; Reverse transcriptase (EC 2.7.7.49)].

GN POL.  
OS Hamster intracisternal A-particle (IAP-H18).  
OC Viruses; Retroviral viruses; Retroviridae; Intracisternal A-particles.  
OX NCBI\_TaxID=11752;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=85264989; PubMed=2991563;  
RA Ono M., Toh H., Miyata T., Awaya T.;  
RT "Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes.";  
RT J. Virol. 55:387-394(1985).  
RL -1 MISCELLANEOUS: READTHROUGH OF THREE TERMINATORS MAY OCCUR: TAA BETWEEN CODONS ATT FOR 660-ILE AND AAA FOR 661-LYS, TAG BETWEEN CCC FOR 859-PRO AND ATT FOR 860-ILE.

CC PIR: A03964; GNHYH.  
CC CCC FOR 859-PRO AND ATT FOR 860-ILE.  
DR HSP: P03366; 1HMV.  
DR InterPro: IPR001037; Integrase\_C.  
DR InterPro: IPR003308; Integrase\_zn.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR000477; RYase.  
DR InterPro: IPR001584; Rye.  
DR Pfam: PF00552; Integrase; 1.  
DR Pfam: PF02022; Integrase\_zn; 1.  
DR Pfam: PF00075; RNaseH; 1.  
DR Pfam: PF00665; Rve; 1.  
DR Pfam: PF00078; Rvt; 1.  
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;  
KW Endonuclease; Polypeptide.  
SQ SEQUENCE 863 AA; 97036 MW; 704AEF7C6B1D886 CRC64;

## Query Match

Best Local Similarity 20.18; Pred. No. 15; Mismatches 110; Indels 121; Gaps 19;

OY 23 KSGAYRHVTRQALPELYEANPLVQSLVITDRDASDADMDADLAGLSPSPVSYNMRYT 82  
DB 420 KTGAGAYVVKRVRISKVQVNEISPOVECLIVLE-----VLEAPGP-----LNIVS 465

OY 83 TTHGIVALKNPVCLTDA-----RRPILNARVEQGLVDGDSYGHRTKNP-- 135  
DB 466 DSSYVV-----NAVNLLEIAGIRSSSRVANIPOKIOALLN-----RFPVEI 509

OY 136 --STAHATLGP-----ADALYELRL-----AHTDEIHALPEAGNPRRYRTSYVG 181  
DB 510 THVRHSGILPGWISGLNDLADAKATLVATALSTHQAQKEFH-----KRHVTAETLR 562

OY 182 RNVTL-----FDTTRMAY-RAVRHSWGPPVAEMHTV-----F 214  
DB 563 RFAALSREAREITYQCQNCCEFLTPPHMGINPRIR-----PLQMKMDVTHIPSFGL 617

OY 215 EIHILNETIITADEFATGCPGLNFKLHSRISRM-VKRN-----FTPETFR--A 261  
DB 618 QYHVSVDTCSGVMFATPLTGKASYVIOHCLEAMSAMGKPRIKTDNGPAVTSQKROFC 677

OY 262 ROKAISLKGASGKGEKGGKGI-----ASGSRRAH--TROQFL 299  
DB 678 ROMDVT-----HLTGLPYNPOGQIGIVERAHRHTLASYL 709

## RESULT 7

ASPX\_PAPHA STANDARD; PRT; 285 AA.

AC 006990;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Acrosomal protein SP-10 precursor (acrosomal vesicle protein-1).  
GN ACV1.  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.  
OX NCBI\_TaxID=9557;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=93183503; PubMed=8442953;  
RA Freemanman A.J., Wright R.M., Flickinger C.J., Herr J.C.;  
RT "Cloning and sequencing of baboon and cynomolgus monkey intra-acrosomal protein SP-10: homology with human SP-10 and a mouse sperm antigen (MSA-63).";  
RT Mol. Reprod. Dev. 34:140-148(1993).  
RN [2]

RP CHARACTERIZATION.  
RX MEDLINE=94220578; PubMed=8167233;  
RA Freemanman A.J., Wright R.M., Flickinger C.J., Herr J.C.;  
RT "Tissue specificity of the acrosomal protein SP-10: a contraceptive vaccine candidate molecule.";  
RL Biol. Reprod. 50:615-621(1994).  
CC -1 SUBCELLULAR LOCATION: NASCENT ACROSOMAL VESICLE OF GOLGI PHASE SPERMATIDS.  
CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1 TISSUE SPECIFICITY: TESTIS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch)).  
CC -----

DR EMBL: S56458; AAB25692.1; -;  
DR EMBL: S56458; AAB25693.1; -;  
KW Signal; Glycoprotein; Repeat; Sperm; Alternative splicing.  
FT SIGNAL 1 21  
FT CHAIN 22 285  
FT DOMAIN 66 95  
FT REPEAT 66 70  
FT REPEAT 71 75  
FT 1-1.  
FT 1-2.

```
FT REPEAT 91 95 1-3.
FT DOMAIN 85 184 7 X 5 AA REPEATS OF S-G-E-H-[AL].
FT REPEAT 85 89 2-1.
FT REPEAT 120 124 2-2.
FT REPEAT 130 134 2-3.
FT REPEAT 145 149 2-4.
FT REPEAT 155 159 2-5.
FT REPEAT 170 174 2-6.
FT REPEAT 180 184 2-7.
FT DOMAIN 110 194 9 X 5 AA REPEATS OF [SV]-G-E-Q-[PSA].
FT REPEAT 110 114 3-1.
FT REPEAT 115 119 3-2.
FT REPEAT 125 129 3-3.
FT REPEAT 135 139 3-4.
FT REPEAT 150 154 3-5.
FT REPEAT 160 164 3-6.
FT REPEAT 165 169 3-7.
FT REPEAT 175 179 3-8.
FT REPEAT 190 194 3-9.
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 172 205 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 285 AA: 30128 MW: 524E115F72AAD0E CRC64;
```

Query Match 5.2%; Score 83; DB 1; Length 285;  
Best Local Similarity 29.9%; Pred. No. 4.1;  
Matches 41; Conservative 11; Mismatches 55; Indels 30; Gaps 7;

```
OY 110 AREOGSLCVLGGDASVYGRITKPNLSTAHATLWGPAD--ALVELRALHTLDEIHALPE 167
DB 16 ANGTSGQSDBSSG--SIDQTSYQQLSGEFLSLENSDAALVETYSGLTTLSE-HGSSR 72
OY 168 AGNPRNRTSTVGRNVLTEDTRMAYRAVRHSGPVPV---AEWEHTVFEH----- 216
DB 73 HSRHTVAEHFGE-----HAESHAEGEPATGAEGEHVTGDEPSCGRPS 120
OY 217 -IHLNETIADFPANG 232
DB 121 GEHLSEGSILG-EHAGS 136

RESULT 8
PTRB_ECOLI STANDARD; PRT; 686 AA.
ID P24555; P78068;
DI 01-MAR-1992 (Rel. 21, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease II (EC 3.4.21.83) (Oligopeptidase B).
GN PTRB OR TLP OR B1845.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OY NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=HB101;
RC MEDLINE=92121142; PubMed=1769955;
RA Kanatani A., Masuda T., Shimoda T., Misoka F., Lin X.S.,
RA Yoshimoto T., Tsuru D.;
RA "Protease II from Escherichia coli: sequencing and expression of the
RT enzyme gene and characterization of the expressed enzyme.";
RL J. Biochem. 110:315-320(1991).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horinchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF LYSYL
AND ARGININYL RESIDUES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of Arg-|-xaa and Lys-|-xaa bonds in
oligopeptides, even when p1' residue is proline.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 59A; ALSO KNOWN AS THE
PROLYL OLIGOPEPTIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL: D10976; BAA01750.1; -.
DR EMBL: AE000278; AAC74915.1; -.
DR EMBL: D90827; BAA15651.1; -.
DR PIR: JQ1151; JQ1151.
DR MEROPS: S09.010; -.
DR Ecocore: EG11004; P1trB.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR004106; Peptidase_S9_N.
DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR002470; Proligo_prase.
DR Pfam: PF00326; Peptidase_S9.1.
DR Pfam: PF02897; Peptidase_S9_N.1.
DR PRINTS: PR00862; PROLIGOPTASE.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolyase; Serine protease; Complete proteome.
FT ACT_SITE 532 532 CHARGE RELAY SYSTEM.
FT ACT_SITE 617 617 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 652 652 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 682 686 ATPAD -> LRLTKYFPDQVSVLNAAPSCCPGY (IN
FT REF. 1).
SQ SEQUENCE 686 AA: 79490 MW: B41AD388044D56B2 CRC64;
```

Query Match 5.2%; Score 83; DB 1; Length 686;  
Best Local Similarity 20.2%; Pred. No. 12;  
Matches 60; Conservative 38; Mismatches 139; Indels 60; Gaps 10;

```
OY 5 ETLPESWLPKRPLASAEKSGAYRHYTORALEP-----YTEANPLVQSLVITDRDA 58
DB 101 QSAFSEEMDEWETLLDANKRAAHSEFYSMGMAITPNTIMALAEDELSRQGIPTRL 160
OY 59 SDADWADIALGDPSPSYVSNRVTYTGHIYALKNPVCLTDAARRRPINILAREOGLCD 118
DB 161 ETGNWYPELLDNEPSEYVWAN---DSWIFYV-----RKHPVTLLP--YQVWRH 204
OY 119 VLGGDASVYGRITKPNLSTAHATLWGPADALVELRALH-TLDEIHAL-----PEAG 169
DB 205 AIGTPASQDKLIYEKDDIYVSLHKTTSKHVYVILHSAFTSEVRLLAEMADAEFPV 264
OY 170 NPNRN-----YTRSTVGRNVLTEDTRMAYRAVRHSGPVAEWEHTVFE 215
DB 265 LPRRKHEYSIDHYQHRFYLRSNRHKNGKNGLY-RTNR-----RDOQWELIPPRENMLE 319
OY 216 HIHLNETIADFPANGPGLNLEIKLSRSI-----SRWVWNTPTPTFFRAR 262
```

Db	320	GTTTCTDWLVVEESRGGLTSLRQIRNRKTRREVIGIAFDPAVYTWIAVNEPETARLR	376		
<hr/>					
RESULT 9					
POL_IPMAI	ID	STANDARD:	PRT:	814 AA.	
AC	P12894;				
Dt	01-OCT-1989	(Rel. 12, Created)			
Dt	01-OCT-1989	(Rel. 12, Last sequence update)			
Dt	01-NOV-1995	(Rel. 32, Last annotation update)			
De	Probable POL polypeptide.				
OS	Mouse Intracisternal A-particle (IAP-IL3).				
OC	Viruses; Retroid viruses; Retroviridae; Intracisternal A-particles.				
Ox	MCHI_taxid=11754;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86286596;	PubMed=3016667;			
RA	Ymer S., Tucker W.O.J., Campbell H.D., Young I.G.;				
RT	"Nucleotide sequence of the intracisternal A-particle genome inserted				
RL	5' to the interleukin-3 gene of the leukemia cell line WEHI-3B."				
CC	-1- MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on ways				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-slb.ch/announce/				
CC	or send an email to license@isb-slb.ch).				
CC	-----				
DR	EMBL; X04120; CAAT2732.1;				
DR	PIR; A23597; GNMSTP.				
DR	HSSP; P04585; IPT1.				
DR	InterPro: IPR001037; Integrase_C.				
DR	InterPro: IPR003308; Integrase_zn.				
DR	InterPro: IPR002156; RNaseH.				
DR	InterPro: IPR004777; RYase.				
DR	InterPro: IPR001584; Rve.				
DR	Pfam; PF00552; Integrase_1.				
DR	Pfam; PF02022; Integrase_zn_1.				
DR	Pfam; PF00075; rnaaseh; 1.				
DR	Pfam; PF00665; rve; 1.				
KW	Pfam; PF00078; rvt; 1.				
KW	Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;				
KW	Endonuclease; Polypeptide.				
SQ	SEQUENCE 814 AA; 91000 MW; A369620A450F729A CRC64;				
<hr/>					
Query Match	5.2%;	Score 82.5;	DB 1;	Length 814;	
Best Local Similarity	19.0%;	Pred. NO. 17;			
Matches 54;	Conservative 44;	Mismatches 119;	Indels 67;	Gaps 11.	
OY	7 LEPESMLRRKPL-----ASAEGSAYRHVRYORALEPLYEIANPLVMQSIVTD----	55			
Db	363 VEPQLTV-RNPLKNGIVYYTDSKTIQAGIYVANGKVSVNQYNENSPQVECLVLEVLT	421			
OY	56 -----RDMSADDMAA-----DLALPSPSYSMNRRVTGTHGVYALKNPVCULTDAARR	104			
Db	422 FLKPINTVISDYSVVNAVNLLVAVGVIPRSSRVANIFOOIOIYVLSSRRSPVYITH----	476			
OY	105 PINLARVEOGICDVLDGSDASYGHRTKNPNLSTANTATLGMPADALYEURLAHNTIDEHA	164			
Db	477 -----VRAHSG----LPPMALGNLDADKATVVAALASSPEEARNH-----NNFHV	521			
OY	165 LPKAGNPRNRYTRSTVGNNVTLFTDTDMA-----YRAVRHSWGGRVAEMERTV---	213			
Db	522 TAEFLRSFFSLTKRKARDIVITOCSCCFELPVPHGVINDPGIR-----PLVQMWDVTHV	576			
OY	214 -----FEHHLNERTIIDEF-----TGPLGUNELKHLSRSISRW 249				
Db	577 SSFGKLQYLHWSIDTCSCIMFPASPLTGKAKASHVIOTHCLEAMSAW 620				

RESULT	10				
MURF_MACTU	STANDARD:	PRT:	510 AA.		
ID	MURF_MACTU				
AC	006220;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	UDP-N-acetylmutramoyl-alanyl-D-glutaryl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)				
DE	(D-alanyl-D-alanine-adding enzyme).				
GN	MURF OR RV2157C OR MT2216 OR MTCy270.11.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RC	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikita A.,				
RA	Bisbal W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RL	Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.				
CC	-I- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL				
CC	STEP IN THE SYNTHESIS OF UDP-N-ACETYL-MURAMYL-PENTAPEPTIDE, THE				
CC	PRECURSOR OF MUREIN (BY SIMILARITY).				
CC	-I- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmutramoyl-L-alanyl-D-				
CC	glutaryl-meso-2,6-diaminopimelate + D-alanyl-D-alanine = ADP				
CC	+ phosphate + UDP-N-acetylmutramoyl-L-alanyl-D-glutamyl-6-				
CC	carboxyl-L-lysyl-D-alanyl-D-alanine.				
CC	-I- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.				
CC	-I- SUBUNIT: MONOMER (BY SIMILARITY).				
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).				
CC	-I- SIMILARITY: BELONGS TO THE MURCEB FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; Z95388; CAB08670.1; -				
DR	EMBL; AE007068; AAK46500.1; -				
DR	HSSP; P11880; IGc4.				
DR	TIGR; MT2216; -				
DR	TubercuList; RV2157c; -				
DR	InterPro; IPR000713; Mur_ligase.				
DR	InterPro; IPR0004101; Mur_ligase.C.				
DR	Pfam; PF02125; Mur_ligase; 1.				
DR	Pfam; PF02875; Mur_ligase.C; 1.				
DR	Peptidoglycan synthesis; Cell division; Cell wall; Ligase;				
DR	Peptidoglycan synthesis; Cell division; Cell wall; Ligase;				

KM ATP-binding; Complete proteome.  
 NP\_BIND 136 142 ATP (POTENTIAL).  
 FT CONFLICT 49 49 A -> R (IN REF. 2).  
 FT CONFLICT 442 442 G -> E (IN REF. 2).  
 SQ SEQUENCE 510 AA; 51632 MW; 7BEB3EDD1BA2EC4 CRC64;

Query Match 5.1%; Score 82; DB 1; Length 510;  
 Best Local Similarity 26.1%; Pred. No. 10;  
 Matches 80; Conservative 24; Mismatches 119; Indels 84; Gaps 19;

QY 28 RHVTRQALAEPLYEANPLVITDRDASADNADLAGLPSPEYGMNRTTGH 87  
 DB 224 REVIAQTKALP--QAVP--HSGAVYLN--ADDPAAVAAKTLAARVRSR-DNTGDV 275  
 QY 88 VYALKNPVCTDARRR-----PINLARVEQGLGDVGDASYGHRITKKNPLSTAHATLW 143  
 DB 276 ---WAGVSLDELARPRFTLHAHDAQAEVRLGVC---GD---HOVT-NALCAAAYAL- 322  
 QY 144 GPADALYELRALAHITDELHALPEAGNPRNRVTRSTVGRNVTLFDTTRMAY----- 195  
 DB 323 -ECGASVEQVAALT-----AAPVSRHRMQVT--TRGDGVTVIDD---AYNANPDSMR 370  
 QY 196 -----RAVHSMG--GPVAMEHTVEFHHILNETITADEFATGGLNDELK 240  
 DB 371 AGIQAIAWIAHOPEATRRSMAVLGEVME-----LGEPAIAEHDRIGRLAVR-- 416  
 QY 241 HLSRSISRWVRNFTPEFPAROKAISLRGASKGKGEKHK-----GIASGASRAHT 294  
 DB 417 ---LDVSRLLVVG--TGHSISAMHGAVALLEGASGSEATADHGADRTAVNVADGDAALAL 472  
 QY 295 RQGFLEG 301  
 DB 473 RAELRPG 479

RESULT 11  
 PKP3\_MOUSE STANDARD; PRT; 797 AA.

AC 090Y23;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Plakophilin 3.  
 GN PKP3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99310799; PubMed=10381383;  
 RA Bonne S., van Hengel J., Nollet F., Kools P., van Roy F.,  
 RT "Plakophilin-3, a novel armadillo-like protein present in nuclei and  
 RT desmosomes of epithelial cells.";  
 RL J. Cell Sci. 112:2265-2276(1999).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN JUNCTIONAL PLAQUES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes.  
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 8 ARM REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF136719; AAD55892.1; -  
 DR MGD: MGI:1891830; Pkpk.  
 DR InterPro: IPR000225; Armadillo.  
 DR Pfam: PF00514; Armadillo\_seg; 5.

DR SMART; SK00185; ARM; 3.  
 DR PROSITE; PS50176; ARM\_REPEAT; 1.  
 KM Cell adhesion; Cytoskeleton; structural protein; Nuclear protein;  
 KM Repeat.  
 FT REPEAT 305 348 ARM 1.  
 FT REPEAT 351 390 ARM 2.  
 FT REPEAT 393 432 ARM 3.  
 FT REPEAT 449 487 ARM 4.  
 FT REPEAT 491 536 ARM 5.  
 FT REPEAT 596 637 ARM 6.  
 FT REPEAT 645 684 ARM 7.  
 FT REPEAT 689 730 ARM 8.  
 SQ SEQUENCE 797 AA; 87298 MW; A9DB0A76C79C126 CRC64;

Query Match 5.1%; Score 82; DB 1; Length 797;  
 Best Local Similarity 22.9%; Pred. No. 18;  
 Matches 69; Conservative 36; Mismatches 124; Indels 72; Gaps 16;

QY 9 PESWLPKPKPLASAEKSGAYRHVTRQALAEPLYEANP-----LYWQSL----- 51  
 DB 153 PTPMPTRPVSEFHRGGAASRADYD-TLSLPSRLGPGGLDRYSVSEGLEPAAASTYR 211  
 QY 52 -VITDRASDA-----DNADLAGLPS-----PSYSMNRVTT-----TGHYYAL 91  
 DB 212 AYAKRQASGSSSRAGLDPEATEGPPRTIRAPAMKTLQRFOSHSRSGTSGVSGAG 271  
 QY 92 KNPVCLDAAARRRINILARVEG-LGDVGDASYGHRITKKNPLSTAHATMGPADALY 150  
 DB 272 LEPAARAPSVRSLSL---ADSGHLPVGRGLDSTYGR-TLQRLSSGPFDDIDIPSAVKY 327  
 QY 151 -----ELRALAHITDELHALPEAGNPRNRVTRSTVGRNVTLFDTTRMAYRAVHSMG 204  
 DB 328 LMASDPMLQYGAAYIGHRCSDAAAKQAASGLAVRLVLFTHNQ---EYGRHATGA 384  
 QY 205 -----PVAMEHTVEFHHILNETITADEF--ATGPL-GLNELKHSNIS 247  
 DB 385 MRNLIDNVNKLALVE-ENGIFELRLTLRQO--DDELKKNVTGILNWLSSSHLKDRLA 441  
 QY 248 R 248  
 DB 442 R 442

RESULT 12

RADA\_PSEAE STANDARD; PRT; 453 AA.

AC P96963;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA repair protein rada homolog (DNA repair protein sms homolog).  
 GN RADA OR SMS OR PA4609.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRDI;  
 RA Howell M.L., Heur M., Klotz M.G., Hassett D.J.;  
 RT "Pseudomonas aeruginosa oxidative stress operon";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.U., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;



RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen." ;  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REPAIR OF ENDOGENOUS ALKYLATION DAMAGE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RADA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@sib-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U89384; AAB9466.1; -.  
 DR EMBL; AE004875; AAC07997.1; -.  
 DR InterPro; IPR001984; Lon\_endopep.  
 DR InterPro; IPR001553; RecA.  
 DR PRINTS; PR00830; ENDOLAPTASE.  
 DR SMART; SM00382; AAA.1.  
 DR PROSITE; PS50162; RACA.2; 1.  
 KW DNA repeat; ATP-binding; zinc-finger; DNA-binding; Complete proteome.  
 FT ZN\_FING 10 27 C4-TYPE (POTENTIAL).  
 FT NP\_BIND 96 103 ATP (POTENTIAL).  
 FT CONFLICT 142 142 E -> K (IN REF. 1).  
 FT CONFLICT 240 240 E -> D (IN REF. 1).  
 FT CONFLICT 243 244 GR -> AA (IN REF. 1).  
 FT CONFLICT 247 247 L -> M (IN REF. 1).  
 FT CONFLICT 258 258 V -> F (IN REF. 1).  
 FT SEQUENCE 453 AA; 48411 MW; AAD23FEFAIEED5 CRC64;  
 SO  
 Query Match 5.1%; Score 81.5; DB 1; Length 453;  
 Best Local Similarity 22.1%; Pred. No. 10; Mismatches 62; Conservative 32; Indels 95; Gaps 13;  
 Matches 62; Conservative 32; Mismatches 91; Indels 95; Gaps 13;  
 QY 1 MDSEFTLPESWLPKRP-----LASAEKSGA---YHVTROALPLEPYEA 43  
 Db 170 IDSIQITFEQ-IGSAPGVAQVRESAAMLVKAGSGAIFLVGHVREGAL-----A 222  
 QY 44 NPLVMSQIVTD-RDASDADMAADLAGLPSPSYSMNRVTTHGIVYALK--NPVCLTD 99  
 Db 223 GPRVLEHWVDVLYFEQSGDGRLLRLRAVKNFAGVNEVGFMGMDKGIKEVNSPSAI-- 280  
 QY 100 AARRPIMLLARVEQGLDVLGDSYGHRTKNPLSTAHATLMGRADALYELRALHTL 159  
 Db 281 -----FLTRAGPAAV-----PGSYVMATWESGRPLAEVQALVDT- 314  
 QY 160 DEIHALPEAGNPFRNVTSTYGRNLT-EDTTRMAYRAVRHSHGCPVAEWEHTVFEHT- 217  
 Db 315 -----SHLAMPRR-----VTGLGDQRLMLLAVLHRHGG-IPTYQDVFLANV 357  
 QY 218 -----HLNFTIADPERATGPLGIN 237  
 Db 358 GGVKYLETASDLALMAAVSSILRNRLPHDLVGEVGLS 397  
 RESULT 13  
 SPCQ\_HUMAN STANDARD; PRT; 2564 AA.  
 AC Q9H254; Q9HCD0; Q9H3G8; Q9H1K7; Q9H1K8; Q9H1K9;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)  
 DE (beta-IV spectrin).  
 GN SPTBN4 OR SPTBN3 OR KIAA1642.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21316449; PubMed=11294830;  
 RA Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L.,  
 RA Gwynn B., Peters L.L., Lux S.E.;  
 RT "A new spectrin, beta-IV, has a major truncated isoform that associates with promyelocytic leukemia protein nuclear bodies and the nuclear matrix."  
 RT J. Biol. Chem. 276:23974-23985(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4).  
 RX MEDLINE=20539976; PubMed=11086001;  
 RA Beggs S., Aguiar D., Dirks R. Jr., Maksimova E., Stabach P.,  
 RA Hermet J.-M., Zhang J.-P., Philbrick W., Stepien V., Ort T.,  
 RA Solimena M.;  
 RT "BetaIV spectrin, a new spectrin localized at axon initial segments and nodes of Ranvier in the central and peripheral nervous system."  
 RT J. Cell Biol. 151:985-1002(2000).  
 RN [3]  
 RP SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).  
 RC TISSUE-Brain;  
 RX MEDLINE=20450683; PubMed=1097877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."  
 RL DNA Res. 7:273-281(2000).  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic islets.  
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@sib-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF311855; AAC2473.1; -.  
 DR EMBL; AF082075; AAC38874.1; -.  
 DR EMBL; AY004226; AAF93171.1; -.  
 DR EMBL; AY004226; AAF93172.1; -.  
 DR EMBL; AY004227; AAF93173.1; -.  
 DR EMBL; AB046862; BAB13468.1; -.  
 DR MIM; 606214; -.  
 DR InterPro; IPR001589; Actinin\_act\_bind.  
 DR InterPro; IPR001715; Calponin\_hom.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR002017; Spectrin.  
 DR InterPro; IPR001605; Spectrin\_PH.  
 DR Pfam; PF00307; CH\_2.  
 DR Pfam; PF00169; PH\_1.  
 DR Pfam; PF00435; spectrin\_18.  
 DR PRINTS; PRO0683; SPECTRINPH.  
 DR SMART; SM00033; CH\_2.  
 DR SMART; SM00233; PH\_1.  
 DR SMART; SM00150; SPEC; 16.  
 DR PROSITE; PS00019; ACTININ\_1; 1.  
 DR PROSITE; PS00020; ACTININ\_2; 1.  
 DR PROSITE; PS50021; CH\_2.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein; Alternative splicing.  
 FT DOMAIN 1 282 ACTIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 61 165 CH 1.  
 FT DOMAIN 180 282 CH 2.  
 FT REPEAT 309 354 SPECTRIN 1.



```

FT REPEAT 398 419 SPECTRIN 2.
FT REPEAT 429 533 SPECTRIN 3.
FT REPEAT 535 642 SPECTRIN 4.
FT REPEAT 644 771 SPECTRIN 5.
FT REPEAT 773 879 SPECTRIN 6.
FT REPEAT 881 985 SPECTRIN 7.
FT REPEAT 1019 1086 SPECTRIN 8.
FT REPEAT 1088 1197 SPECTRIN 9.
FT REPEAT 1199 1303 SPECTRIN 10.
FT REPEAT 1305 1408 SPECTRIN 11.
FT REPEAT 1410 1513 SPECTRIN 12.
FT REPEAT 1515 1619 SPECTRIN 13.
FT REPEAT 1621 1725 SPECTRIN 14.
FT REPEAT 1727 1832 SPECTRIN 15.
FT REPEAT 1834 1940 SPECTRIN 16.
FT REPEAT 1942 2046 SPECTRIN 17.
FT REPEAT 2048 2107 SPECTRIN 18.
FT DOMAIN 2418 2527 PH.
FT VARSPLIC 1 1257 MISSING (IN ISOFORM 3).
FT VARSPLIC 1258 1286 AVQAAGLLRQGNITGEQAQAVTLLER -> MPHYPSCS
FT VARSPLIC 1287 1309 SAPSLGTIPQIQIQLAHNR (IN ISOFORM 3).
FT VARSPLIC 1310 2564 NOENLRQAQOMKLDLELOH -> CLIIHPALHPWE
FT VARSPLIC 2113 2154 PPIYLRSS (IN ISOFORM 2).
FT VARSPLIC 2113 2154 MISSING (IN ISOFORM 2).
FT VARSPLIC 2113 2154 TERIKRQSKPPPTLLKRFEDPTLAKAKAPILPGCY
FT VARSPLIC 2155 2564 E -> PRREDHNPQVQDPQWQHTKPSLPKPKRANEKTAR
FT CONFLICT 604 608 MISSING (IN ISOFORM 4).
FT CONFLICT 714 714 MISSING (IN ISOFORM 4).
FT CONFLICT 1189 1189 L -> S (IN REF. 2).
FT CONFLICT 1193 1193 E -> K (IN REF. 2).
FT CONFLICT 1193 1193 E -> K (IN REF. 2).
FT CONFLICT 1331 1331 G -> S (IN REF. 1).
SQ SEQUENCE 2564 AA; 288982 MW; 52CDE7D11D601ECC CRC64;

```

Query Match 5.1%; Score 81.5; DB 1; Length 2564;  
 Best Local Similarity 22.3%; Pred. No. 91;  
 Matches 65; Conservative 38; Mismatches 113; Indels 75; Gaps 14;

```

QY 11 SWL-PKRPPLASAEKSG-----AYRHYTORALFPIEIANPLMOSIVITDRDASDMDA 64
DB 1529 AWQERPLIAMOTERNGIAVOOHITKNOGLRREIOAGPRLEEV-----ERA 1578
QY 65 ADLAGLPSPSYVSMNVTTHGIVALKNFVCLTDARRRP--INLARYEQGLCDVLAG 122
DB 1579 GALASLRSPAEAVRGLDGLQASAMA-----GLREAERQVLDAAFYEQYFFDVAEV 1633
QY 123 DASYGRIT-----KNPLST-----AHATLMGPADALYELRALAHITDETHALPEAG 169
DB 1634 EAWLGEQELLMSEDKGKDEQSTLQKLKHLQLEQGVEN-YE-ESTIQLSRCCRLLEMG 1691
QY 170 NP-----RRNTRSTVGRNVTLFDTRRMAYRAVHSMGSPVAEMHTV--- 213
DB 1692 HPDSEISRROSOVDRLVYALKELGEBRRVALFOQYTWLYOLSRO-----VSELEHWTAEK 1746
QY 214 -----FEHILNLETTIADFEA--TGPLGLNELKHLSSISRWV 250
DB 1747 EYVAGSPELGDFEHVSVLOERF--SEFASSETMAGSERLAANVQWDELI 1795
RESULT 14
AMPK_METH
ID AMPK_METH STANDARD; PRT; 299 AA.
AC 027355;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Methionine aminopeptidase (EC 3.4.11.18) (MAP) (Peptidase M).
GN MAP OR MTH1296.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.

```

```

OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H:
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE
CC + PEPTIDE.
CC -1- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C; ALSO KNOWN AS THE
CC MAP FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000894; AAB85776.1; -.
DR HSSP; P56218; 1XGS.
DR MERO; M24.001; -.
DR InterPro: IPR002468; MAP_2.
DR InterPro: IPR001714; Methano_Prase.
DR InterPro: IPR000994; Peptidase_M24.
DR Pfam; PF00557; Peptidase_M24; 1.
DR PRINTS; PR00599; MAPEPTIDASE.
DR PROSITE; PS01202; MAP_2; 1.
KW Hydrolyase; Aminopeptidase; Cobalt; Complete proteome.
FT METAL 84 84 COBALT 2 (BY SIMILARITY).
FT METAL 95 95 COBALT 1 AND 2 (BY SIMILARITY).
FT METAL 158 158 COBALT 1 (BY SIMILARITY).
FT METAL 191 191 COBALT 1 (BY SIMILARITY).
FT METAL 285 285 COBALT 1 AND 2 (BY SIMILARITY).
SQ SEQUENCE 299 AA; 33431 MW; DA37006A0638167 CRC64;

```

Query Match 5.0%; Score 80.5; DB 1; Length 299;  
 Best Local Similarity 27.2%; Pred. No. 7.2;  
 Matches 25; Conservative 9; Mismatches 33; Indels 25; Gaps 3;

```

QY 130 ITNPSTAHATLMGPADALYELRALAHITDETHALPEAGN-----RRNTRSTVG 181
DB 213 LRSPRLVHAR-----RVLCRIREYHALPFAQRMLEEYDARLNASMR--- 258
QY 182 RNVTLFDTTRMAYRAVHSMGSPVAEMHTV 213
DB 259 ---MLIOSRAIYPHYLREKSGANVAMQWERTV 287

```

```

RESULT 15
ENTC_ECOLI
ID ENTC_ECOLI STANDARD; PRT; 391 AA.
AC P10377; P77099;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Isochorismate synthase entC (EC 5.4.99.6).
GN ENTC OR B0593 OR Z0735 OR ECS0632.
OS Escherichia coli, and
OS Escherichia coli O157:H7.

```

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562, 83334;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-K12;  
 CC MEDLINE=90236256; PubMed=2110093;  
 CC Elkins M.F., Earhart C.E.;  
 CC "Optically factor from group A streptococci is an apoproteinase.";  
 CC FEMS Microbiol. Lett. 56:35-40(1988).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-K12;  
 CC MEDLINE=89123153; PubMed=2536681;  
 CC Ozenberger B.A., Brickman T.J., McIntosh M.A.;  
 CC "Nucleotide sequence of Escherichia coli isochorismate synthetase  
 CC gene and evolutionary relationship of isochorismate synthetase  
 CC and other chorismate-utilizing enzymes.";  
 CC J. Bacteriol. 171:775-783(1989).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-K12 / MG1655;  
 CC MEDLINE=97426617; PubMed=9278503;  
 CC Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 CC Mau B., Shao Y.;  
 CC "The complete genome sequence of Escherichia coli K-12.";  
 CC Science 277:1453-1474(1997).  
 CC [4]  
 CC SEQUENCE FROM N.A.  
 CC Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 CC Federapfel N., Hyman R., Kaiman S., Komp C., Kuri O., Lew H., Lin D.,  
 CC Natch A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 CC Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC [5]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 CC MEDLINE=21074935; PubMed=11206551;  
 CC Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 CC Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 CC Pofal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 CC Grobbeck E.J., Davis N.W., Lin A., Dimalanta E.T., Potamocis K.,  
 CC Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 CC Welch R.A., Blattner F.R.;  
 CC "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 CC Nature 409:529-533(2001).  
 CC [6]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-O157:H7 / RIMD 0509952;  
 CC MEDLINE=21136231; PubMed=11236796;  
 CC Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 CC Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 CC Iida T., Takami H., Honda T., Sasaki K., Ogawara N., Yasunaga T.,  
 CC Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 CC "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 CC O157:H7 and genomic comparison with a laboratory strain K-12.";  
 CC DNA Res. 8:11-22(2001).  
 CC [7]  
 CC SEQUENCE OF 1-33 FROM N.A.  
 CC MEDLINE=90230305; PubMed=2139473;  
 CC Brickman T.J., Ozenberger B.A., McIntosh M.A.;  
 CC "Regulation of divergent transcription from the iron-responsive fepB-  
 CC entC promoter-operator regions in Escherichia coli.";  
 CC J. Mol. Biol. 212:669-682(1990).  
 CC [8]  
 CC PARTIAL SEQUENCE AND CHARACTERIZATION.  
 CC MEDLINE=90241936; PubMed=2139795;  
 CC Lin J., Quinn N., Berchold G.A., Walsh C.T.;  
 CC "Overexpression, purification, and characterization of isochorismate  
 CC synthase (EntC), the first enzyme involved in the biosynthesis of  
 CC enterobactin from chorismate.";  
 CC Biochemistry 29:1417-1425(1990).

CC -1- CATALYTIC ACTIVITY: Chorismate = Isochorismate.  
 CC -1- PATHWAY: ENTEROBACTIN BIOSYNTHESIS. ENTEROBACTIN IS AN IRON-  
 CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE  
 CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- INDUCTION: By iron starvation.  
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 CC TRPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; M24142; AA16100.1; -;  
 CC DR EMBL; M36700; AA18491.1; -;  
 CC DR EMBL; AE000165; AAC73694.1; -;  
 CC DR EMBL; U82598; AAB40793.1; ALT\_INIT.  
 CC DR EMBL; AE005239; AAG54928.1; -;  
 CC DR EMBL; AP002552; BAB34055.1; -;  
 CC DR EMBL; X53274; CAA37371.1; -;  
 CC DR PIR; J70497; SYECIK.  
 CC DR PIR; S04323; S04323.  
 CC DR ECGene; EG10261; entC.  
 CC DR InterPro; IPR000350; Chorismate\_bind.  
 CC DR Pfam; PF00425; chorismate\_bind.1.  
 CC DR ProDom; PD000779; Chorismate\_bind.1.  
 CC DR Enterobactin biosynthesis; Iron transport; Isomerase;  
 CC Complete proteome.  
 CC KW CONFLICT 305 306 SG -> TA (IN REF. 1).  
 CC FT SEQUENCE 391 AA; 42931 MW; 62882569DFC1AC4 CRC64;

Query Match 5.0%; Score 80.5; DB 1; Length 391;  
 Best Local Similarity 22.2%; Pred. NO. 10;  
 Matches 61; Conservative 36; Mismatches 89; Indels 89; Gaps 13;

QY	9	PESWPRKPLASAEKSGRYHYTORALEPTEENP-----LVMSIVTDDASD	60
DB	91	PESW---QSFSGOEKQASARRRFTNSQSLNVEBROAIPQPTPEQWVARAAATATATQYDK	147
QY	61	-----ADMAADLAGLPSPSYVSMRVYTTGHIYA-----LK-----	92
DB	148	VYLSRLIDITDDAID-SGVLLERLIAQNPVSYNHFVLAOGVLLGASPELLLRKQDER	206
QY	93	-NPVCLTAAARRPINLAR-----VEGGLCDVLCGDA5YGRITKN	133
DB	207	FSSISLAGSARQPDVEDRENGNRLASKEKRHEHELVTOAMKEVLERSESLH-VPSS	265
QY	134	PLSTAHATLM-----CPADALYELRALAHITDEIHLPEGNGRRVNTSTVGRANTL	186
DB	266	POLITPTLMLATPEFGKANSQENALTACILHPTPLD--SGFPQATO--VIALELP	321
QY	187	EDTFRMAYRAVRHSGVGV-----AEWEHTV	213
DB	322	FD-----RELFGIVGWCDESGENGWYTI	346

Search completed: October 4, 2002, 15:36:31  
 Job time: 203 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 46.38 Seconds

(without alignments)  
1130.174 Million cell updates/sec

Title: US-09-720-583a-2

Perfect score: 1596  
Sequence: 1 MDSFETLFPESWLPKRPLAS.....IAGSASRRHTRQDFLEGLS 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_tvirus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeoplasmid:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1596	100.0	303	2	09F6J9
2	403.5	25.3	295	2	09F5N6
3	343	21.5	341	2	09ZEX6
4	338.5	21.2	310	2	030683
5	329	20.6	336	2	052148
6	322.5	20.2	308	2	051635
7	321.5	20.1	311	2	051631
8	320.5	20.1	308	2	051636
9	316.5	19.8	308	2	051639
10	311.5	19.5	306	2	09L457
11	306	19.2	302	2	09L457
12	304	19.0	310	2	045279
13	303.5	19.0	311	2	051632
14	303	19.0	311	2	051630
15	298	18.7	297	2	09L458
16	289.5	18.1	304	2	051637

17	286.5	18.0	304	2	051638	051638 escherichia
18	283	17.7	313	2	09K124	09K124 edwardsiella
19	282.5	17.7	297	2	051629	051629 shigella sp
20	282.5	17.7	304	2	051633	051633 escherichia
21	276.5	17.3	317	2	P97024	P97024 brevibacter
22	275.5	17.3	296	2	051634	051634 shigella so
23	268.5	16.8	320	2	09F4V6	09F4V6 brevibacter
24	266	16.7	296	2	09XBN1	09XBN1 rhodococcus
25	259.5	16.3	306	2	045286	045286 bifidobacter
26	243.5	15.3	289	2	050999	050999 nelsisera g
27	198	12.4	247	2	P94637	P94637 corynebacte
28	198	12.4	437	2	050474	050474 pseudomonas
29	191.5	12.0	437	2	09Z3S9	09Z3S9 pseudomonas
30	189.5	11.9	278	2	09X743	09X743 halomonas e
31	189	11.8	437	2	093TF3	093TF3 pseudomonas
32	180	11.3	320	2	056299	056299 thiothacillu
33	146	9.1	157	2	031107	031107 pseudomonas
34	117	7.3	524	2	0938U2	0938U2 corynebacte
35	108.5	6.8	524	2	093T64	093T64 corynebacte
36	107.5	6.7	521	2	093MK3	093MK3 corynebacte
37	105	6.6	84	2	046074	046074 corynebacte
38	99.5	6.2	2164	2	093N86	093N86 streptomyce
39	98	6.1	387	11	099JU4	099JU4 mus musculu
40	98	6.1	693	11	093UC9	093UC9 mus musculu
41	98	6.1	1461	16	005819	005819 mycobacteri
42	97	6.1	3332	5	09VEF8	09VEF8 drosophila
43	93.5	5.9	228	17	09YD65	09YD65 aeropyrum p
44	93	5.8	355	12	09IGZ5	09IGZ5 maize strea
45	92	5.8	355	12	091MG6	091MG6 maize strea

## ALIGNMENTS

RESULT 1	
09F6J9	PRELIMINARY; PRT; 303 AA.
AC 09F6J9;	
DT 01-MAR-2001 (TREMBLrel. 16, Created)	
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE REPI.	
OS Propionibacterium freudenreichii.	
OG Plasmid p545.	
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC Actinomycetales; Propionibacteriaceae; Propionibacteriaceae;	
OC Propionibacterium.	
OX NCBI_TaxID=1744;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-LMG16545;	
RX MEDLINE-21091935; PubMed-11157209;	
RA Jore J.P.M., van Luitk N., Luiten R.G.M., van der Werf M.J.,	
RA Pouwels P.H.;	
RT "Efficient transformation system for Propionibacterium freudenreichii	
RT based on a novel vector.";	
RL Appl. Environ. Microbiol. 67:499-503(2001).	
DR EMBL: AF291751; MAG25294.1; -.	
DR InterPro: IPR004322; Replicase.	
DR Pfam: PF03090; Replicase; 1.	
KW Plasmid.	
SQ SEQUENCE 303 AA; 33613 MW; 5E6503BFA70A1AD7 CRC64;	

Query Match 100.0%; Score 1596; DB 2; Length 303;  
Best Local Similarity 100.0%; Pred. No. 8.1e-133;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1	MDSFETLFPESWLPKRPLASKSGAYRHTYRRALELPYIEENPLVMOSLVTTDDASD 60
DB 1	MDSFETLFPESWLPKRPLASKSGAYRHTYRRALELPYIEENPLVMOSLVTTDDASD 60
OY 61	ADMAADLAGLSPFSYVSMNRVTTGHIYVALKNPVCLTDAARRRPTNLARVEGLCDVL 120

```

Db 61 ADMADLAGLSPSVSNRRNTTGHITVYALKNPVCLDARRRINILARVEGLCYLV 120
Qy 121 GGDASYGHRITKKNPLSTAHATLWGPADALYELRALAHTLDEIHALPEAGNRRNVTRSTV 180
Db 121 GGDASYGHRITKKNPLSTAHATLWGPADALYELRALAHTLDEIHALPEAGNRRNVTRSTV 180
Qy 181 GRANTLPETTHMAVYRAVRHSMGPGVAEMHTVEPHILHNETTIADDFATGPGIGLNLK 240
Db 181 GRANTLPETTHMAVYRAVRHSMGPGVAEMHTVEPHILHNETTIADDFATGPGIGLNLK 240
Qy 241 HLSRISRWVRNFTPEFRAROKAISLRGASKGEGKHGIGIASGASRRARHTPOCLE 300
Db 241 HLSRISRWVRNFTPEFRAROKAISLRGASKGEGKHGIGIASGASRRARHTPOCLE 300
Qy 301 GLS 303
Db 301 GLS 303

```

```

RESULT 2
Q9F5N6 PRELIMINARY; PRT; 295 AA.
ID Q9F5N6 AC Q9F5N6;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
DE REPLICATION PROTEIN.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP RYabchenko L.E., Novikov A.D., Golyshin P.N., Yankenko A.S.;
RA "Rhodococcus erythropolis" plasmid pN30 putative replicase (ORF1), and
RT putative DNA-binding replication protein (ORF2).";
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF312210; AAC29854.1; -.
KW Plasmid.
SQ SEQUENCE 295 AA; 33678 MW; 05E9197C2FD6DB6 CRC64;

```

Query Match 25.3%; Score 403.5; DB 2; Length 295;  
 Best Local Similarity 36.0%; Pred. No. 1.4e-27;  
 Matches 104; Conservative 49; Mismatches 101; Indels 35; Gaps 11;

```

Qy 13 LPRKPLASAE-KSGAVRHTRQRALELPYIRANPLVMQSLVITRDASADMAA--DLAG 69
Db 20 LPHRPFATNNIQROQYR-MSRDALAMRYVHSHALIGSTIVICDHDVDAAMRAFEQPSD 78
Qy 70 LPSPSVSMNRVYTTGHIYVAL-KNPVCLDARRRINILARVEGLCYLVGGDASYG 128
Db 79 HPADNVAOS-PSGRAHGWLGNHVCRTDSARLTPLRYAHRLETGKISVGGEFAVG 137
Qy 129 RITKNPLSTAHATLWGPADALYELRALA--HTLDEIHALPEAGNRRNVTRSTVGRNTL 186
Db 138 QLTNNPILHPDWEITYGPATP-YTLRQLATITPPROM-----PRRDRAVGLGRNTM 188
Qy 187 FDTTMMAYRAVRHSMGPGVAEMHTVEPHILHNETTIADDFATGPGIGLNLKHSRI 246
Db 189 FDAIRRAAYRPMQHRNGTGRDMDHLVIOCHAVN-----TEFTT-PLPFTVEVRAIQA 242
Qy 243 SKWIMRNTEYOYARAQAHL-----GQKGGKATILAQEAVR 279

```

```

RESULT 3
Q9ZEX6 PRELIMINARY; PRT; 341 AA.
ID Q9ZEX6 AC Q9ZEX6;

```

```

DT 01-MAY-1999 (Tremblrel, 10, Created)
DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE REPA PROTEIN.
GN REPA.
OS Thibacillus sp.
OC Bacteria; Proteobacteria; beta subdivision; Thiodacillus.
OX NCBI_TaxID=924;
RN [1]
RP SEQUENCE FROM N.A.
RA Perera J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ007958; GAA07792.1; -.
DR InterPro; IPR004322; Replicase.
DR Pfam; PF03090; Replicase; 1.
KW Plasmid.
SQ SEQUENCE 341 AA; 38283 MW; E0F1B8A7532F55AD CRC64;

```

Query Match 21.5%; Score 343; DB 2; Length 341;  
 Best Local Similarity 32.1%; Pred. No. 3.6e-22;  
 Matches 98; Conservative 40; Mismatches 129; Indels 38; Gaps 11;

```

Qy 13 LPRKPLASAEKSGAVRHTRQRALELPYIRANPLVMQSLVITDD--ASDADW--AADL 67
Db 19 LPHRPFATNNIQROQYR-MSRDALAMRYVHSHALIGSTIVICDHDVDAAMRAFEQPSD 78
Qy 68 AGLPSPSVSMNRVYTTGHIYVAL-KNPVCLDARRRINILARVEGLCYLVGGDASYG 127
Db 79 VGLPPNIAVNNHRRKRLHLLYLEAGVCRTLEHLKRLRYIASTERYTAVLGDPOYA 138
Qy 128 HRTKNPLSTAHATLWGPADALYELRALAHTLDEIHA---LPEAGNRRNVTRSTVGR 182
Db 139 GLICKNPENDR---W---QVWEIHGNPYTLAEIAAYVDLTAKAARLPNSASSEFGLGR 190
Qy 183 NVTLEDTTMMAYRAVRHSMG-GPVAEMHTVEPHILHNETTIADDFATGPGIGLNLK 241
Db 191 NCTMFDTRFMAVYGAVRYPMPNGIDRWSKAVLELQAIN---GEPPO-PLPFAEVKA 244
Qy 242 LRSRISRWVRNFTPEFRAROKAISLRGASKGEGKHGIGIASGASRR 291
Db 245 TGKISGWTWRNMP---AGLDLIERHTPEKQERKATNQADIAVMGIGIASGAR 301
Qy 292 AHTRQ 296
Db 302 LSRGQ 306

```

```

RESULT 4
Q30683 PRELIMINARY; PRT; 310 AA.
ID Q30683 AC Q30683;
DT 01-JAN-1998 (Tremblrel, 05, Created)
DT 01-JAN-1998 (Tremblrel, 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE PUTATIVE REPLICATION PROTEIN.
GN REPA.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-NI86/21;
RC MEDLINE=98015402; PubMed=9353918;
RA De Mot R., Nagy I., De Schrijver A., Pattanapipitpaisal P.,
RA Schoofs G., Vanderleyden J.;
RT "Structural analysis of the 6-kb cryptic plasmid pFAJ2600 from
RT Rhodococcus erythropolis NI86/21 and construction of Escherichia coli-
RL Microbiology 143:3137-3147(1997).

```



Db	179	LFEMRKAAVRAIRGWM--PVFSQWLDVAIQVEMVNSLAPV-----PLSPAECRAIGK	230
QY	245	SISRWYMNENPETFR-----ARQAIISLRGSKSGKEGHR---GGIASGA	288
Db	231	SIATKTHNFPETFAQYVADTHTPETPEIQAAARGKRGKIGGASRKGAVATSA	282
RESULT	7		
ID	051631	PRELIMINARY:	PRT: 311 AA.
AC	051631.		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	REP PROTEIN.		
OS	Escherichia coli.		
OG	Plasmid COLE2-GEI602.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95050306; PubMed=7525540;		
RA	Hiraga S., Sugiyama T., Itoh T.,		
RT	"Comparative analysis of the replicon regions of eleven COLE2-related		
RT	plasmids."		
RL	J. Bacteriol. 176:7233-7243(1994).		
DR	EMBL: D310017. BAA06295.1; .		
DR	InterPro: IPR000792; HTH LuxR.		
DR	InterPro: IPR004322; Replicase.		
DR	InterPro: IPR001440; TRP.		
DR	Plam; Pf03090; Replicase; 1.		
KW	plasmid.		
SO	SEQUENCE 311 AA; 35081 MW; F91655855E321C2A CRC64;		
Query Match	20.1%; Score 321.5; DB 2; Length 311;		
Best Local Similarity	31.3%; Pred. No. 2.5e-20;		
Matches	94; Conservative 53; Mismatches 124; Indels 29; Gaps 12		
QY	10	ESWLPRKPLASAEKSGAVRHVTRQALEPLPYEANPLVQMSQSVI--TDRDASDADAAADL	67
Db	9	EENLPHPRYHDDLAFLGRISGKGRALLARYIQGNQPHQFPLVLDVDEGAALDMSDN	68
QY	68	AGLSPSPSVSNRRVTTGHIYALKNPVCLDIAARRRPINLIARVEQGLCDVYGDASTG	127
Db	69	A--PAPNTVKNPVNGAHLILYALINAVTAPDASVKALKYAAAEALRACERKIGADVNTS	126
QY	128	HRITKPNLSTAHATLWGPADALYELRALAHNTLDEI--HALPEAGNPRNRYTRS--TVGRNVT	185
Db	127	GLICNNPHE-----W-----LVMEKREAVYTIDELADLYDLSASARSRIDKHGMGRNCT	178
QY	186	LEPDTTMAAYRAVNRHSGWGPV-AEMEHTVFEHIIHLNETIINADFEATGRLGNETKHLNR	244
Db	179	LEEMTKRKAAYRAIRGWM--PVFSQWLDVAIQVEMVNSLAPV-----PLSPAECRAIGK	230
QY	245	SISRWYMNENPETFRAROKAISLSGASGKGEHGKGIASGA---SRKATTRQDFLEG	301
Db	231	SIATKTHNFPETFAQYVADTHTPETPEIQAAARGKRGKIGGASRKGAVATSA	282
RESULT	8		
ID	051636	PRELIMINARY:	PRT: 308 AA.
AC	051636.		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	REP PROTEIN.		
OS	Shigella sonnei.		
OG	Plasmid COLE2-CT14.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Shigella.		

Query Match	Best Local Similarity	Matches	90; Conservative	19.8%; Score 316.5; DB 2; Length 308; Best Local Similarity 30.9%; Pred. No. 6.9e-20; Matches 90; Conservative 48; Mismatches 124; Indels 29; Gaps 10;
NCBI_Taxid=624;				
SEQUENCE FROM N.A.				
MEDLINE=95050306; PubMed=7525540;				
Hiraga S., Sugiyama T., Itoh T.,				
"Comparative analysis of the replicon regions of eleven <i>ColE2</i> -related				
plasmids."				
J. Bacteriol. 176:7233-7243(1994).				
EMBL: D30061; BAA06299.1; -				
InterPro: IPR004322; Replicase.				
InterPro: IPR001440; TPR.				
Pfam: PF03090; Replicase; 1.				
Plasmid.				
SEQUENCE 308 AA; 34922 MW; 55863D1650C03EAE CRC64;				
Query Match	20.1%; Score 320.5; DB 2; Length 308;			
Best Local Similarity	32.2%; Pred. No. 3e-20;			
Matches	94; Conservative 45; Mismatches 122; Indels 31; Gaps 11;			
ESMPPRKPLAESAEGSAVRYTRQALPELPEYEAFLPVMSQSLVIT--TPRDSADAMADL 67				
9 EENLPHRYHDDLAFGLRISGKALARIYIQNQPHAQFVLVDIDREGALIDMSDRN 68				
AGLPSPSVSMNRVYTTGHIYVALKNPVCITDAAARRPINLLARVEQGLCDVLGDSAYG 127				
69 A--PAPNTIVKNPVGSHALLLYALNIAVRTAPDASVRAKFAAAIECALCEKIGADVNS 126				
128 HRIKRNPLSTAHATLMGPDALYELRALHTLDEI-HALPEAGNRRRVTRS-TVGRNVT 185				
127 GLICKENPHLE---W---LVMERKEAVYTLDELADYLDLSASRSRISDKHYGGRNC 178				
186 LFDTRMAYAVRAVRSWVGPy-AEMEHTVFEHILNLTIIADEFAFGPLGINELKHSR 244				
179 LEEMTRKWAYAIRGQW--PVFSQWLDVAVIQREVENYSLPV-----PLSPAECRAIGK 230				
245 SISRWARNFPEPFR-----AROKAISLRGASKGKGGKHK--GGIASGA 288				
231 SIATYTHRNFPETFAOYVADTHPEIQANGRGKIGKAKSKRGAVATSA 282				
RESULT 9				
051639	PRELIMINARY;	PRF;	308 AA.	
AC 051639;				
DT 01-NOV-1996 (TREMBlrel. 01, Created)				
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)				
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)				
DE REP PROTEIN.				
OS Escherichia coli.				
OC Plasmid ColE2-J.				
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC Escherichia.				
OX NCBI_Taxid=562;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=95050306; PubMed=7525540;				
RA Hiraga S., Sugiyama T., Itoh T.,				
RT "Comparative analysis of the replicon regions of eleven <i>ColE2</i> -related				
plasmids."				
RL J. Bacteriol. 176:7233-7243(1994).				
EMBL: D30061; BAA06302.1; -				
InterPro: IPR004322; Replicase.				
DR InterPro: IPR001440; TPR.				
DR Pfam: PF03090; Replicase; 1.				
DR Plasmid.				
SO SEQUENCE 308 AA; 34851 MW; 2245A0EAD06FBF085 CRC64;				





```

RESULT 12
ID 045279 PRELIMINARY; PRT: 310 AA.
AC 045279;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PLASMID PRBL1 POTATIVE THERA REPLICASE (ORF310).
OS Brevibacterium 11ens.
OC Plasmid autonomous circular plasmid PRBL1.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Brevibacteriaceae; Brevibacterium.
OX NCBI_TaxID=1703;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RBL;
RX MEDLINE=97092385; PubMed=6938050;
RA Ankri S., Bouvier I., Reyes O., Predali F., Leblon G.;
RT "A Brevibacterium 11ens PRBL1 replicon functional in Corynebacterium
RT glutamicum."
RL Plasmid 36:36-41(1996).
DR EMBL: U39878; AAB03568.1; -.
DR InterPro: IPR004322; Replicase.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ
SEQUENCE 310 AA; 35266 MW; 30B4DF1F26A98BF4 CRC64;

Query Match 19.0%; Score 304; DB 2; Length 310;
Best Local Similarity 32.8%; Pred. No. 8.8e-19;
Matches 106; Conservative 34; Mismatches 137; Indels 46; Gaps 13;

QY 3 SEFTLPESWLPKPKLASAE-KSGAYRHVTRORALELPYIEANPLYMOSLVTTRDASDA 61
DB 5 STET-WGGMWLPMPPLASDLDLGGIYR-TSRNNALELRIENPOSTLSNLVVDIDHDA 62
QY 62 -----DNAADLAGLPSPSYVMNRVTGTHIVALKNEVCITDARRRPINLLARV 112
DB 63 LMRAMNRRAM-----QPNAYVENPANGAHAVMALAEFVTEYARRKPLAYAAV 114
QY 113 EOGLCDVYGDSYGHRTKKNPLSTAHNTLWGPADALYELRALHTLDEIHALPRAG--- 169
DB 115 TEGLRKSYDGKGYSGLTITKNPTHDQWASW-LTDHLNLTBELTEHLVSDMPDQWQR 173
QY 170 NPRNRVSTVGKRVNLTEDTRMAVRAVRHSGPVAEWEHTVEHILINETII----- 225
DB 174 TKRKNPV--GKGRNCTLETVRMVDYVARPI---RKRNHPEDRDHLEALIVNLCQ 227
QY 226 -ADEPATGPLGNEKLHSRISRKVMNRFT---PEFFRARKAISLRGASGKGECH 280
DB 228 GGNSTFSEALPASEIRATIRSFYKWTITRYTGWLDRTSOSKSAHYHRT---GRKGL 284
QY 281 KGIASGASRAHTRQOFLGIS 303
DB 285 KGVVSGQV---RQKALERS 303

RESULT 13
ID 051632 PRELIMINARY; PRT: 311 AA.
AC 051632;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE REP PROTEIN.
OS Escherichia coli.
OC Plasmid COLE2Imm-K317.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven COLE2-related
RT plasmids."
RL J. Bacteriol. 176:7233-7243(1994).
DR EMBL: D30058; BAA06296.1; -.
DR InterPro: IPR000792; HTH_LuxR.
DR InterPro: IPR004322; Replicase.
DR InterPro: IPR001440; TPR.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ
SEQUENCE 311 AA; 35370 MW; 5569BA824F2B2646 CRC64;

Query Match 19.0%; Score 303.5; DB 2; Length 311;
Best Local Similarity 30.3%; Pred. No. 9.8e-19;
Matches 89; Conservative 53; Mismatches 125; Indels 27; Gaps 10;

QY 10 ESWLPKPKLASAEKSGAYRHVTRORALELPYIEANPLYMOSLVT--TRDASDADMAADL 67
DB 9 EENLPHRPHYHTDDLAFGRISGKGRALLARYIQOQPHAQFWLVDPVREGAALIDMSDRN 68
QY 68 AGLPSPSYVMNRVTGTHIVYALKNPVCLTDAAARRRPINLLAREOGCLCVLGSASYG 127
DB 69 A-PAPNITVKNPVNGAHHLIALNIAVTRAPDASVKALKTAAYERSLCEKLCADAVYS 126
QY 128 HRTKNPLSTAHNTLWGPADALYELRALHTLDEI-HALPENGPRNRVTRS-TVGRNVT 185
DB 127 GLICKNPFHLE---W---LVMEWREAYITLDELADYLDLSASRRSIDHNGGRCH 178
QY 186 LFDTRMAVRAVRHSGPVAEWEHTVEHILINETIADDEFATGPLGNEKLHSRIS 245
DB 179 LFEMTRKMAVYAIRQGWPA-FSQWIDAVIQREMYNASLPV-----PLSPPECAIKS 231
QY 246 ISRWVNRNFTETFEAROKAISLRGASGKGECHGKGIASGASRATROFL 299
DB 232 IAKYTHRNFTPETF---AQYVADHTPTLOAKRGKGIANG---EAYVDRFM 279

RESULT 14
ID 051630 PRELIMINARY; PRT: 311 AA.
AC 051630;
RT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE REP PROTEIN.
OS Escherichia coli.
OC Plasmid COLE2-CA42.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050306; PubMed=7525540;
RA Hiraga S., Sugiyama T., Itoh T.;
RT "Comparative analysis of the replicon regions of eleven COLE2-related
RT plasmids."
RL J. Bacteriol. 176:7233-7243(1994).
DR EMBL: D30056; BAA06294.1; -.
DR InterPro: IPR000792; HTH_LuxR.
DR InterPro: IPR004322; Replicase.
DR Pfam: PF03090; Replicase; 1.
KW plasmid.
SQ
SEQUENCE 311 AA; 35298 MW; F2E75BECA0DBA137 CRC64;

Query Match 19.0%; Score 303; DB 2; Length 311;
Best Local Similarity 30.4%; Pred. No. 1.1e-18;
Matches 90; Conservative 55; Mismatches 123; Indels 28; Gaps 12;

QY 8 FPESWLPKPKLASAEKSGAYRHVTRORALELPYIEANPLYMOSLVT--TRDASDADMAA 65

```

DB 8 FDEN-LPHRPHHTDGLAFLRISGKALLARYIQONQPHQFWLVPDVBEGALIDW-S 65  
 QY 66 DLADGPSYVSMNRVTTHGHVYALKNPVCLDPAARRPINLARVEOGICDVLGDAS 125  
 DB 66 DL-NAPAPNITIKNPVGNHLLYALNIAVTPADASVKALKYAAIERALCELTGADV 124  
 QY 126 YGHRITKPNPLSTAHATMGADALYELRALAHILDEI-HALPEAGNRRNVTNS-TVGRN 183  
 DB 125 YSGIICKNPHELE---W---QVMEWEKEAYTLDELADYLDLSTARSIDKHYGMGRN 176  
 QY 184 VTLEDDTRMAVYRAVRHSHMGVPAWEHTEVFEHILNLTIIADEFATGPGLNELKHS 243  
 DB 177 CHLEFEMTRKWAYRAIRGOWPA-FSQWLDAYIQVEMYNASLPV-----PLSPPECRAIG 229  
 QY 244 RSISRWWYNNETPETEFAROKAISLRGASKGEGHKGIGASGSRRAHTROQFL 299  
 DB 230 KSIATKTHRRFTAETF---AQYVADHTPEIQAKRGKRGGIAGK---EAYDDKRFM 279

RESULT 15  
 Q9RLE8 PRELIMINARY; PRT; 297 AA.  
 ID Q9RLE8  
 AC Q9RLE8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE PUTATIVE REPLICON PROTEIN.  
 GN REP.  
 OS Plesiomonas shigelloides (Aeromonas shigelloides).  
 OC plasmid p1184.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Plesiomonas.  
 OX NCBI\_TaxID=703;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O36:H34;  
 RX MEDLINE=21223402; PubMed=11322823;  
 RA Avison M.B., Walsh T.R., Bennett P.M.;  
 RT "PUB060: a broad host-range, DNA polymerase-I-independent ColE2-like  
 RT plasmid."  
 RL plasmid 45:88-100(2001).  
 DR EMBL; AJ249644; CAB56518.1; -;  
 DR InterPro; IPR004322; Replicase.  
 DR Pfam; PF03090; Replicase; 1.  
 KW plasmid.  
 SQ SEQUENCE 297 AA; 33820 MM; FBDDEF09EEFBCE2 CRC64;

Query Match 18.7%; Score 298; DB 2; Length 297;  
 Best Local Similarity 33.4%; Pred. No. 2.8e-18;  
 Matches 101; Conservative 33; Mismatches 138; Indels 30; Gaps 12;

QY 6 TLEPESLTPRKPLASAEKSGAIVRYTORALELPIYIAN-PLVMQSLYT-TDRDASDADW 63  
 DB 7 TLEFNDRLPHKPYFSDLOFGVIRAGKERALLKYYIOFNQPHAMWYLCFVDVDRAGAIDW 65  
 QY 64 AADLAGLPSYSVSMNRVTTHGHVYALKNPVCLDPAARRPINLARVEOGICDVLGD 123  
 DB 66 -ADL-GAPAPTLITIKNDNGHAHLIYALNIAVTPADGRGLKYAAIERALCELTGAD 123  
 QY 124 ASYGHRTKPNPLSTAH--ATLMGPADALYELRALAHILDEIHALPEAGNRRNVTSTVG 181  
 DB 124 AGYSGIICKNP-NHLHMQITVMQP--ELYTIDMLADYLD---LGAANDREILPDVGLG 175  
 QY 182 RNTLEPPTTMAVYRAVRHSHMGVPAWEHTEVFEHILNLTIIADEFATGPGLNELKH 241  
 DB 176 RNCLEPDKTRKWAYRAIRGOW-PEYSQWLOACIERARAYNLQFSA-----PLDENWVG 228  
 QY 242 LSRSSIRWV-----WRNFTPETFRAROKAISLRGASKGEGHKGIGASGSRRAH 293  
 DB 229 IAKSISKMWVTVYRSLGDFDEVYKLTSPVQAYNGRRSKGGGRPSIGEPWLALGISRSY 288  
 QY 294 TR 295

DB 289 FR 290

Search completed: October 4, 2002, 15:36:07  
 Job time: 179 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:08 ; Search time 20.69 Seconds  
(without alignments)  
357.707 Million cell updates/sec

Title: US-09-720-583A-2

Perfect score: 1596  
Sequence: 1 MDSFETLPESWLPKPLAS.....IAGSARRAHTRQOFILEGLS 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	5.3	4302	3	US-08-658-136-5
2	84.5	5.3	4303	2	US-08-460-751-2
3	83.5	5.2	450	4	US-09-191-879-2
4	83.5	5.2	1043	3	US-08-928-361B-30
5	83.5	5.2	1721	3	US-08-700-651-5
6	83.5	5.2	1721	3	US-08-928-361B-6
7	83	5.2	285	1	US-08-292-045-5
8	78.5	4.9	1745	2	US-09-031-485-33
9	78.5	4.9	1745	2	US-08-847-429A-33
10	78.5	4.9	1745	2	US-09-065-474-33
11	78	4.9	973	3	US-08-904-452-4
12	77	4.8	319	2	US-08-683-908-1
13	77	4.8	1711	2	US-08-342-930-2
14	77	4.8	2539	4	US-09-413-814-42
15	77	4.8	3739	4	US-09-320-878-2
16	77	4.8	3739	4	US-09-105-537-33
17	77	4.8	11877	4	US-09-105-537-6
18	76.5	4.8	746	1	US-08-476-519-11
19	76.5	4.8	746	5	PCT-US95-09323-11
20	76.5	4.8	777	1	US-08-476-519-2
21	76.5	4.8	777	5	PCT-US95-09323-2
22	76	4.8	265	1	US-08-292-045-2
23	76	4.8	285	1	US-08-292-045-10
24	76	4.8	685	3	US-08-947-965-74
25	75.5	4.7	501	3	US-08-906-791-2
26	75.5	4.7	501	4	US-09-111-730-1
27	75.5	4.7	773	3	US-08-564-264-1

28	75	4.7	250	4	US-09-216-295-14	Sequence 14, Appl
29	75	4.7	378	4	US-09-500-569-10	Sequence 10, Appl
30	75	4.7	867	2	US-08-938-365-2	Sequence 2, Appl
31	75	4.7	1042	3	US-08-928-361B-11	Sequence 11, Appl
32	75	4.7	1837	3	US-08-928-361B-5	Sequence 5, Appl
33	74.5	4.7	554	4	US-08-943-714-2	Sequence 2, Appl
34	74	4.6	285	1	US-08-292-045-7	Sequence 9, Appl
35	74	4.6	417	2	US-08-672-564-9	Sequence 9, Appl
36	74	4.6	580	2	US-08-672-564-1	Sequence 20, Appl
37	73.5	4.6	417	4	US-08-867-611-20	Sequence 25, Appl
38	73.5	4.6	177	5	PCT-US92-06965A-25	Sequence 9, Appl
39	72.5	4.5	177	2	US-08-737-825-9	Sequence 32, Appl
40	72.5	4.5	606	2	US-08-577-492-32	Sequence 7, Appl
41	72.5	4.5	606	4	US-09-079-630-32	Sequence 2, Appl
42	72.5	4.5	828	3	US-08-481-435-2	Sequence 10, Appl
43	72.5	4.5	850	3	US-08-481-435-10	Sequence 18, Appl
44	72.5	4.5	964	3	US-08-860-339-18	Sequence 5, Appl
45	72.5	4.5	1611	2	US-08-804-227C-5	

#### ALIGNMENTS

RESULT 1  
US-08-658-136-5  
; Sequence 5, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,136  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400  
; TELEFAX: 508-872-5415  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-658-136-5

Query Match 53%; Score 84.5; DB 3; Length 4302;  
Best Local Similarity 21.1%; Pred. No. 26;  
Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps 9;

```

QY      4 FETLEPSPWPRPRLPPLASAEKSGVRRH-----TQRLALEPYLEANDPLWQSLVI 53
Db      1391 FVLQGDSEAML-----VACAMPPEPFRRTWDFGEELAPFTARPEVTFILRDP-----GSLIV 1443

QY      54 TDRASADMAADLACLPSPSYISMAKRVTYTTGTHIYALKNPVCLTDAARRPILINLARVE 113
Db      1444 TVTASNNISANOSALAEVQEPALVTSIKVNSLGLQOPLTFSAVSGREASY----- 1498

QY      114 QGLDVLGVGDASGYHRTTKNPPLSTAHATL-----WGPADALYELRALAHITLD 160
Db      1499 ---LMDLDGGMLEGPETVTHAYNTSGDFYRVAVAGMNEVSEKMLNLYKKRRKGLV----- 1552

QY      161 EIALPPEAGNRRNRVTRSYVGNVTLFDITRMAYRAVRHSW-----GGPVAEME 210
Db      1553 -----WASRTVYPLNGSVSFSTSLSEAGSDVAYSWLCLDRCTPIPGCPTIYS- 1599

QY      211 HTVPEHILNLEPILINDEFRTGPLGINELKHLSRSTRSRV 250
Db      1600 ---TRSYGTENIIVYAL-----NENGSAODSLETVIV 1628

```

```

1 RESULT 2
2 US-08-460-751-2
3 : Sequence 2, Application US/08460751
4 : Patent No. 5891628
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Reeders, Stephen
8 : APPLICANT: Schneider, Michael
9 : APPLICANT: Gluckmann, Sandra
10 :
11 : TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
12 : TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
13 : NUMBER OF SEQUENCES: 35
14 : CORRESPONDENCE ADDRESS:
15 : ADDRESSEE: Pennie & Edmonds
16 : STREET: 1155 Avenue of the Americas
17 : CITY: New York
18 : STATE: New York
19 : COUNTRY: U.S.A.
20 : ZIP: 10036-2711
21 :
22 : COMPUTER READABLE FORM:
23 : MEDIUM TYPE: Floppy disk
24 : COMPUTER: IBM PC compatible
25 : OPERATING SYSTEM: PC-DOS/MS-DOS
26 : SOFTWARE: PatentIn Release #1.0, Version #1.30
27 : CURRENT APPLICATION DATA:
28 : APPLICATION NUMBER: US/08/460,751
29 : FILING DATE: 02-JUN-1995
30 : CLASSIFICATION: 435
31 : PRIOR APPLICATION DATA:
32 : APPLICATION NUMBER: US 08/413,580
33 : FILING DATE: 03-MAR-1995
34 : ATTORNEY/AGENT INFORMATION:
35 : NAME: Coruzzi, Laura A.
36 : REGISTRATION NUMBER: 30,742
37 : REFERENCE/DOCKET NUMBER: 7638-005
38 : TELECOMMUNICATION INFORMATION:
39 : TELEPHONE: (212) 790-9050
40 : TELEFAX: (212) 869-8864/9741
41 :
42 : TELEX: 66141 PENNIE
43 : INFORMATION FOR SEQ ID NO: 2:
44 : SEQUENCE CHARACTERISTICS:
45 : LENGTH: 4303 amino acids
46 : TYPE: amino acid
47 : TOPOLOGY: unknown
48 : MOLECULE TYPE: protein
49 :
50 : US-08-460-751-2
51 :
52 : Query Match 5.3%; Score 84.5; DB 2; Length 4303;
53 : Best Local Similarity 21.1%; Pred. No. 26;
54 : Matches 59; Conservative 35; Mismatches 111; Indels 75; Gaps:
55 :
56 : 4 FETLPFESWLPKRLPLASAKSGAIVRHV-----TRQRALEPLPIEANPLWMSLVI 53

```

Db	1391	FVLQGDPAWL	-----VACAMPEPFYRTWDFGEEAARFARBPVEVTFIYRDP-----GSYL	1443
QY	54	TDROASADMAADL	AGLPSSVSYSNMRRVTTTHGIYALKNPCLDAAARRPILNAVE	113
Db	1444	TVTASNNISANDS	ALVEOPEVLVSIKNGSLGLEDOOPLFSAVGRPAST-----	1498
QY	114	QGLCDVLGDA	SGYHRTKNPLSTAHATL-----WGPAALYELRALAHPLD	160
Db	1499	--LMDLDGGWLG	PEVTHAYNSTGDFTVAVAGMNEVSRSEAMLVNTVRRVGLV----	1552
QY	161	EIHALPEAGN	RRNVRVTSYGRANVTLFDTRMAYPAVHNS-----GGEVAEME	210
Db	1553	-----VNASTV	YPLNGSVSFSSLEAGSDVXRWVLCDCRCPJPGGPTIS-	1599
QY	211	HTVEFHILN	ERTIADFEFATPGLGINKHLSRISRRV	250
Db	1600	--TRSYGTENI	IVTAL-----NEGSAODSLFVVY	1628

```

RESULT 3
US-09-191-879-2
: Sequence 2, Application US/09191879
: Patent No. 6204014
: GENERAL INFORMATION:
: APPLICANT: May, Earl W.
: APPLICANT: Jaworski, Deborah D.
: APPLICANT: Wang, Min
: APPLICANT: Warren, Richard L.
: APPLICANT: Lennox, Anna L.
: APPLICANT: Ingraham, Karen A.
: APPLICANT: Ray, Jennifer
: APPLICANT: Chalkey, Allison F.
: APPLICANT: Holmes, David J.
: APPLICANT: Mazulla, Marie J.
: TITLE OF INVENTION: Dnab
: FILE REFERENCE: GM10183
: CURRENT APPLICATION NUMBER: US/09/191,879
: CURRENT FILING DATE: 1998-11-13
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: FastSeq for Windows Version 3.0.
: SEQ ID NO 2
: LENGTH: 450
: TYPE: prt
: ORGANISM: Streptococcus pneumoniae
US-09-191-879-2

```

Query Match	5.2%	Score 83.5	DB 4	Length 450
Best Local Similarity	22.2%	Pred. No. 0.97		
Matches	62	Conservative	43	Mismatches 99
			Indels	75
			Gaps	15
OY	40	YIE-----ANPLVMOSLV-TTDR-DASDA-----DMAADLAGLDSPSYV--SMNRY	81	
Db	37	YIESDFEKRYAHLRFQAMVYDLSDEGDAIDATVTRITLDNOGDLNIGSLSTIVEIVNSV	96	
OY	82	TTTGHIVALKNPVCLTGAARRRP-----NLLARVQGLCDVLGSD	123	
Db	97	PTSAAEAYYAK--YAERKMLRLRLAKLTLESYNQAEASQPADDEIIAQAERKLIIDVSENA	154	
OY	124	ASYGRIITKNPLSTAHATILWCPADALYEL-----RALAHTLDEIH-----ALPEA	168	
Db	155	NNSGKFNKRDVYLNLMFGNLEARSQOTTDITIGATGYRDLDDHMTTGGHEBELIILARPAV	214	
OY	169	G-----NPRRNVTRSTYGRNVTLFD-----TTPMAYRAVRHSWG---GPAV--EW	209	
Db	215	GKTAFAALNIAQNI-GTKLDKTYAITSLEMGASIVDRMLAAGIVESHSHIRIGQLTDEEM	273	
OY	210	EHTVEFHHLNETIADFEATGPGVLNDELKILSSISIR	248	
Db	274	QKYYTAQGNLANASIIYIDD--TPGIRITEIRNRSKRLAQ	310	

RESULT 4  
US-08-928-361B-30  
; Sequence 30, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIRSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928, 361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: VERNY, HANA  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480,76-1(HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1678  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1043 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-928-361B-30

Query Match 5.2%; Score 83.5; DB 3; Length 1043;  
Best Local Similarity 25.3%; Pred. No. 3,6;  
Matches 74; Conservative 30; Mismatches 100; Indels 89; Gaps 17;

QY 52 VTDDASD-ADWADLGLPSPSYVSMNRYT-----TTGHIY----- 88  
DB 452 VVDEAKQADKQKGLIVPTNSINKDPTNTQYNTGNTGKVIPIGSL 511  
QY 89 -YALNPVCLTDAAARRPINLARVEQGLCDVLGDSYGHRI--TKNPLSTAHATLWG 144  
DB 512 NVPSENTPOQTEITGKPVDTVG-----LPYDPSGELIIDPATKLPIGSVA----- 559  
QY 145 PADATYELRALHTLDEIHALP---EAGNPRNRVTSYGRNVTLFTTRMANAYRAVRHS 201  
DB 560 -GDEIL-TEVLNITTDVETGLPIDLETGLPRDVSGLPOLPNGTLVDPSENK--KPIPS 614  
QY 202 WGGPVAEMETHVFEHILNLTIDEFATG-PL-----GLNELKHSRSISRWVWNFTP 256  
DB 615 HSGFI--NGTSGEOSH-----EKDPSTGKPLDPTNTLHPDEDSGL-----INP 657  
QY 257 ETPRAROKAISLRGASKG-----GREGHGKGIASGASRRRAHTRROOFLGSL 303  
DB 658 ET-----GDKLQSHSGTFMPVPGKPGENGIM-----TPQIILEALN 696

RESULT 5

US-08-700-651-5  
; Sequence 5, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM  
; TITLE OF INVENTION: INJECTIONS  
; FILE REFERENCE: 480,19-4(HV)  
; CURRENT APPLICATION NUMBER: US/08/700,651B  
; CURRENT FILING DATE: 1997-08-14  
; EARLIER APPLICATION NUMBER: 08/415,751  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1721  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 5.2%; Score 83.5; DB 3; Length 1721;  
Best Local Similarity 25.3%; Pred. No. 8;  
Matches 74; Conservative 30; Mismatches 100; Indels 89; Gaps 17;

QY 52 VTDDASD-ADWADLGLPSPSYVSMNRYT-----TTGHIY----- 88  
DB 1130 VVDEAKQADKQKGLIVPTNSINKDPTNTQYNTGNTGKVIPIGSL 1189  
QY 89 -YALNPVCLTDAAARRPINLARVEQGLCDVLGDSYGHRI--TKNPLSTAHATLWG 144  
DB 1190 NVPSENTPOQTEITGKPVDTVG-----LPYDPSGELIIDPATKLPIGSVA----- 1237  
QY 145 PADATYELRALHTLDEIHALP---EAGNPRNRVTSYGRNVTLFTTRMANAYRAVRHS 201  
DB 1238 -GDEIL-TEVLNITTDVETGLPIDLETGLPRDVSGLPOLPNGTLVDPSENK--KPIPS 1292  
QY 202 WGGPVAEMETHVFEHILNLTIDEFATG-PL-----GLNELKHSRSISRWVWNFTP 256  
DB 1293 HSGFI--NGTSGEOSH-----EKDPSTGKPLDPTNTGLHPDEDSGL-----INP 1335  
QY 257 ETPRAROKAISLRGASKG-----GREGHGKGIASGASRRRAHTRROOFLGSL 303  
DB 1336 ET-----GDKLQSHSGTFMPVPGKPGENGIM-----TPQIILEALN 1374

RESULT 6  
US-08-928-361B-6  
; Sequence 6, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIRSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/928,361B
FILING DATE:  12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US 60/026,062
FILING DATE:  13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME:  Verry, Hana
REGISTRATION NUMBER:  30,518
REFERENCE/DOCKET NUMBER:  450.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE:  650-324-1677
TELEFAX:  650-324-1678
INFORMATION FOR SEQ ID NO:  6:
SEQUENCE CHARACTERISTICS:
LENGTH:  1721 amino acids
TYPE:  amino acid
STRANDEDNESS:
TOPOLOGY:  linear
MOLECULE TYPE:  protein
US-08-928-361B-6

```

Query Match	5.2%	Score 83.5	DB 3	Length 1721
Best Local Similarity	25.3%	Pred. No. 8		
Matches 74	Conservative 30	Mismatches 100	Indels 89	Gaps 17

QY	52	VILDKDASD-ADMAADLADGAPSPSYSSMNRVT-----TTGIV-----	88
Db	1130	VWPEEFAKDADAGKGGGLLPPPNSSINKBPVNTQYSSNTGRIINPEKCVIPGSLP	1189
QY	89	-YAKNNVCCTDARRRRPILILARVQGGCDLVGGGASGYHRI-----TKNP	144
Db	1190	NYPSEFPQOQDELTGCGPVDVTG-----LPYPSGELIDPATKLPJIGSAV	1237
QY	145	PADLYELRLALATLDEIHALP--EAGNPRNVSTVGRAWTEPDTTBMAYAAV	201
Db	1238	-GDBIL-TEVLNTITTEDVIGSLPIDELETGLPRDVSGLPOLPGTILVDPNK--	1293
QY	202	WGGVVAEMETVVEHILHILNETIIADEPANG-PL-----GLNELKLYSR	256
Db	1293	HSQGT---NGTSGQSH-----EKDPSGKPLDPNTGLHPFDEDSGL-----	1335
QY	257	ETFRAROKAISLRGASKG-----GREGGKGGIASGASRAHTRQOFLG	303
Db	1336	ET-----GDKLGGSHGTFMPVPGPQENGIM-----TEQILLEAN	1374

RESULT 7  
 US-08-292-045-5  
 Sequence 5, Application US/08292045  
 Patent No. 5602005  
 GENERAL INFORMATION:  
 APPLICANT: HERR, JOHN C.  
 APPLICANT: WRIGHT, RICHARD M.  
 TITLE OF INVENTION: PRIVATE INTRA-ACROSOMAL SPERM ANTIGEN  
 TITLE OF INVENTION: FOR USE IN A CONTRACEPTIVE VACCINE  
 NUMBER OF SEQUENCES: 29  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. Jefferson Davis Highway, Suite 400  
 City: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

```

1      APPLICATION NUMBER:  US/08/292,045
2      FILING DATE:
3
4      CLASSIFICATION:  424
5
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER:  US 07/318,551
8      FILING DATE:  03-MAR-1989
9      APPLICATION NUMBER:  US 07/481,491
10     FILING DATE:  16-FEB-1990
11     APPLICATION NUMBER:  US 07/858,798
12     FILING DATE:  27-MAR-1992
13     ATTORNEY/AGENT INFORMATION:
14     NAME:  OBLON, NO. 5602005mad F.
15     REGISTRATION NUMBER:  24,618
16     REFERENCE/DOCKET NUMBER:  494-109-0 CITE
17
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE:  (703) 413-3000
20     TELEFAX:  (703) 413-2220
21     TELE:  248655 OPAT UR
22
23     INFORMATION FOR SEQ ID NO:  5:
24
25     SEQUENCE CHARACTERISTICS:
26
27     LENGTH:  285 amino acids
28     TYPE:  amino acid
29     TOPOLOGY:  linear
30
31     MOLECULE TYPE:  protein
32
33     US-08-292-045-5

```

Query Match	5.2%	Score 83;	DB 1;	Length 285;
Best Local Similarity	29.9%	Pred. NO. 0.54;		
Matches 41;	Conservative 11;	Mismatches 55;	Indels 30;	Gaps 7.

[illegible]

RESULT 8  
 US-09-031-485-33  
 Sequence 33, Application US/09031485  
 Patent No. 5824306  
 GENERAL INFORMATION:  
 APPLICANT: Tang, Liang  
 TITLE OF INVENTION: DIPOPTERIN AND BRUGIA ANKRRIN  
 TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 TITLE OF INVENTION: USES THEREOF  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Carol Talkington Verser, Ph.D.  
 ADDRESS: Hesk Corporation  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80525  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/031,485  
 FILING DATE:  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:



APPLICATION NUMBER: US/08/847.429  
 FILING DATE: 24-APR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: HW-5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 33:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1745 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-031-485-33

Query Match 4.9%; Score 78.5; DB 2; Length 1745;  
 Best Local Similarity 22.8%; Pred. No. 30;  
 Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YIEANPLVMQSLVITDRDASD-----ADMAADLA 68  
 DB 1385 AITLPEYTGPEPWSKLFYSEASLTKYVGAFHETAEPPDNLPLAHVALLIGADWHRLAR 1444  
 QY 69 GLPSP-----SYVSNRVTTGHIYALK--NPVCLDAARRRPINLLARVEQGL 116  
 DB 1445 ALEVPDIDRQVRHQLVGLAVTILRIWIFLKKQATPVALSALDRIGRDVVR-EMDR 1503  
 QY 117 CVDLGDAASYGHRITKNPLSTAHATLWGPADALYELRALAHLDLDETHALPEAGNPRRN-- 174  
 DB 1504 AEKLDG-----LKGTPVS--HIS--GPS-----TILSSTLLEY-----AGDRRRHAE 1541  
 QY 175 VTRS-----TVGNVTLFD-----TTRMAYRAVRHSM 202  
 DB 1542 VTMAQORLAQEPFQGVNGTTPGDEPEKQSFHEEEVAVSEIRTVRTERHVDSE 1601  
 QY 203 GGPVAMEHTV--FEHILNLTIIADFPANGPLGLNELKHLRSISHWVNRFTPETF 259  
 DB 1602 NGPIVE-ERTITTYDDVAVNEELVDKIV--PLNEEQEKMDRVN-REVENNF----- 1652  
 QY 260 RAROKAISLRGA---SKGKGEGHKGGIASGASRRAPHTROQFLEG 301  
 DB 1653 --EQETSKEGTFCQCTHEKEKDDGSLKTKMDSHVROIJFFDG 1696

RESULT 9  
 US-08-847-429A-33  
 ; Sequence 33, Application US/08847429A  
 ; Patent No. 5827692  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Liang  
 ; APPLICANT: Blehm, E. Scot  
 ; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; NUMBER OF SEQUENCES: 85  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carol Talkington Verser, Ph.D.  
 ; ADDRESSEE: Heska Corporation  
 ; STREET: 1825 Sharp Point Drive  
 ; CITY: Fort Collins  
 ; STATE: Colorado  
 ; COUNTRY: USA  
 ; ZIP: 80525  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Windows 95  
 ; SOFTWARE: Wordperfect for Windows, Version 7.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/847.429A

FILING DATE: 24-APR-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verser, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: HW-5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/493-7272  
 TELEFAX: 970/484-9505  
 INFORMATION FOR SEQ ID NO: 33:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1745 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-847-429A-33

Query Match 4.9%; Score 78.5; DB 2; Length 1745;  
 Best Local Similarity 22.8%; Pred. No. 30;  
 Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YIEANPLVMQSLVITDRDASD-----ADMAADLA 68  
 DB 1385 AITLPEYTGPEPWSKLFYSEASLTKYVGAFHETAEPPDNLPLAHVALLIGADWHRLAR 1444  
 QY 69 GLPSP-----SYVSNRVTTGHIYALK--NPVCLDAARRRPINLLARVEQGL 116  
 DB 1445 ALEVPDIDRQVRHQLVGLAVTILRIWIFLKKQATPVALSALDRIGRDVVR-EMDR 1503  
 QY 117 CVDLGDAASYGHRITKNPLSTAHATLWGPADALYELRALAHLDLDETHALPEAGNPRRN-- 174  
 DB 1504 AEKLDG-----LKGTPVS--HIS--GPS-----TILSSTLLEY-----AGDRRRHAE 1541  
 QY 175 VTRS-----TVGNVTLFD-----TTRMAYRAVRHSM 202  
 DB 1542 VTMAQORLAQEPFQGVNGTTPGDEPEKQSFHEEEVAVSEIRTVRTERHVDSE 1601  
 QY 203 GGPVAMEHTV--FEHILNLTIIADFPANGPLGLNELKHLRSISHWVNRFTPETF 259  
 DB 1602 NGPIVE-ERTITTYDDVAVNEELVDKIV--PLNEEQEKMDRVN-REVENNF----- 1652  
 QY 260 RAROKAISLRGA---SKGKGEGHKGGIASGASRRAPHTROQFLEG 301  
 DB 1653 --EQETSKEGTFCQCTHEKEKDDGSLKTKMDSHVROIJFFDG 1696

RESULT 10  
 US-09-065-474-33  
 ; Sequence 33, Application US/09065474  
 ; Patent No. 6063599  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Liang  
 ; APPLICANT: Blehm, E. Scot  
 ; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN  
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; NUMBER OF SEQUENCES: 171  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carol Talkington Verser, Ph.D.  
 ; ADDRESSEE: Heska Corporation  
 ; STREET: 1825 Sharp Point Drive  
 ; CITY: Fort Collins  
 ; STATE: Colorado  
 ; COUNTRY: USA  
 ; ZIP: 80525  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Windows 95  
 ; SOFTWARE: Wordperfect for Windows, Version 7.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/065.474

```

; FILING DATE: 24-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37/459
; REFERENCE/DOCKET NUMBER: HW-5-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-065-474-33

Query Match      4.9%; Score 78.5; DB 3; Length 1745;
Best Local Similarity 22.8%; Pred. No. 30;
Matches 79; Conservative 36; Mismatches 118; Indels 113; Gaps 19;

QY 35 ALELP-YTEANPLVMQSLVITDRDASD-----ADMADLA 68
   |||
Db 1385 AILPEPTGEPWVSKRLFTSEASLYKYGAFHETAEPDNPLAHVALLIGADMHRLAR 1444
   |||
QY 69 GLPSP-----SYVSNMNRVTTHGIYVALK---NPVCLTDARRRPINLLARVEGL 116
   |||
Db 1445 ALEVPDIDIRHQLVGLAEVITLRIMFLKREQATPVLRSLQIRGRDQYR--EMDR 1503
   |||
QY 117 CDVIGGASGHRITKPLSTAHNTLWGPADALYELALHTLDEIHALPAGNPRRN-- 174
   |||
Db 1504 AEKLDG-----LEGTPVS--HIS--GPS-----ITLSTLLEV-----AGDRRRHAE 1541
   |||
QY 175 VTRS-----TVGRNVTLFD-----TRMAYRAVVRHSW 202
   |||
Db 1542 VTMAQOHLAGEPFOQVGYNGTBPDEPEPKQSTHEREEVAVSEITVYNTENHVDSE 1601
   |||
QY 203 GGPVMEHTV---FEHILHNETIADFEATGFLGINKLHSRSISRWRNFTPETF 259
   |||
Db 1602 NGPIVE-ERTITTYEDDVAVNEEIVDKTY--PLNEEEDKKMDRV--REVENMF----- 1652
   |||
QY 260 RAKQKALSILGA---SKGKEGKHKGSIASGASRAHTRQOFLLEG 301
   |||
Db 1653 --EQQETSKETGCGQTHERKDDGSLKTYMKDSHVRIQIFPDG 1696

RESULT 11
; US-08-904-452-4
; Sequence 4, Application US/08904452
; Patent No. 6083742
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo M.
; TITLE OF INVENTION: Mammalian Deep Orange Proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,452
; FILING DATE: 31-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
```

```

; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1240.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-904-452-4

Query Match      4.9%; Score 78; DB 3; Length 973;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 49; Conservative 27; Mismatches 106; Indels 44; Gaps 7;

QY 5 ETLPEESMLPRKPLASAEKSGAVRHVTRORALELPYTEANPLVMQSLVITDRDASDADWA 64
   |||
Db 19 QTCGSPVIGIPHSYVSAHLEKEVPIFTKQRYDFTPSRITSIVYSCNQLCMSLGGKDTLLR 78
   |||
QY 65 ADLAGLPSPTVSKNRVT-----TTGHIYVALKNPVCU---TDAARRRPINLLA 110
   |||
Db 79 IDLGKASEPNRVELGRDQDAKVRHMKFLDHTGSHLVALSSTEVLNMRNGOKARP---LA 135
   |||
QY 111 RVEQGLDVGADASYGHRITKNP-----LSTAHNTLWGPADALYELRALA 156
   |||
Db 136 RWKQGLVESVGMKAMKNESSTGPIIVGTAGQITFEALSLASBGLGFPADLT----- 189
   |||
QY 157 HTLDEIHALPEAGNP---RRNVTRSTVGRNVTLFDTTRMAYRAV 198
   |||
Db 190 --FRPLVYNDEGGPAPVCSLEAERGPDRGQFY-ATTROQLFGFI 232

RESULT 12
; US-08-683-908-1
; Sequence 1, Application US/08683908
; Patent No. 5830693
; GENERAL INFORMATION:
; APPLICANT: Shimizu, Sakayu
; APPLICANT: Kobayashi, Michihiko
; TITLE OF INVENTION: A Regulatory Factor Involved in
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,908
; FILING DATE: 19-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185626/1995
; FILING DATE: 21-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Davidson, Clifford M
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 3821004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 768-3800
; TELEFAX: (212) 382-2124
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 319 AMINO ACID RESIDUES  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
ORIGINAL SOURCE:  
ORGANISM: Rhodococcus rhodochrous  
INDIVIDUAL ISOLATE: J1  
CELL TYPE: unicellular organism  
US-08-683-908-1

Query Match 4.8%; Score 77; DB 2; Length 319;  
Best Local Similarity 21.3%; Pred. No. 3;  
Matches 48; Conservative 33; Mismatches 94; Indels 50; Gaps 11;

QY 4 FETLPESW-----LPKRLASAEKSGAVRHVTRORALEPLYEANPLVMOSLVTDRD 57  
DB 114 YDLFTDMRMASVQFPFRYSFGFTD-----FGKMTAVVGGDRGIGRVSSFMSTIND 168  
QY 58 ASDADMAADLAGLPSPSYVSMNRTTGHIVYALKNPVCLTDARRRPINLLARVEQGLC 117  
DB 169 ATDAGLAEEVASLHNSAVDLLSAIRT-----ELADQAAASDGLL---CVLAVIRQNL 220  
QY 118 D-VLGSDASGHRITKNPLSTAH---ATLNGPADALYELRALHTLDETHALPEAGNP 171  
DB 221 DPNLCASQIAAEHNVS---VRTLHRLFSANGGVAEHIRMIR-----LERIKT--ELADP 270  
QY 172 RRVNTRSTVGRNVTLFDTTRMAY-----RAVRHSGGPAVEM 209  
DB 271 -----TSRRYTTISALARKMGFLDPSTFSRAFKDAYGITAREW 307

RESULT 13

US-08-342-930-2  
Sequence 2, Application US/08342930  
Patent No. 5821084

GENERAL INFORMATION:

APPLICANT: OLMSTED, ELIZABETH A.

APPLICANT: MAURO, LAURA J.

APPLICANT: DAVIS, ALAN R.

APPLICANT: DIXON, JACK E.

TITLE OF INVENTION: OSTROBLAST-TESTICULAR PROTEIN TYROSINE

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/342,930

FILING DATE: 21-NOV-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: KONSKI, ANTOINETTE F.

REGISTRATION NUMBER: 34,202

REFERENCE/DOCKET NUMBER: 20344-20975.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1711 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-342-930-2

Query Match 4.8%; Score 77; DB 2; Length 1711;  
Best Local Similarity 20.1%; Pred. No. 42;  
Matches 56; Conservative 39; Mismatches 109; Indels 74; Gaps 11;

QY 20 SAEKSGAVRHVTRORALEPLYEANPLVMOSLVTDDASDADMAADLAGLPSPSYVSMN 79  
DB 580 AALKASWYHPGGGRDAFHLRLRPLTLESEVLPREAQNFSAQLTACCEQVOVIS-- 637  
QY 80 RVTTGHIVYALKNPVCLTDARRRPINLLA---RVEQGLCDVGLGDASYGHRITKNPL 135  
DB 638 --TLMGSESSSANANGWTPPSAPTLVNVTSDAPTQVQSWAHVPGGRSRY--QVTLYOE 693  
QY 136 STAHAT-LMGPADALYELRALHTLDETHALPEAGNRRRVNTRSTVGRNVTLFDTTRMA 194  
DB 694 STRVATSIMGP-----KDDGTSFGLTPTGT-----K 719  
QY 195 YRAVRHSGGP-----VAWEHTYPEHTHLN---ETIADFAPTGPG-----L 236  
DB 720 YKVEVISMAGPLTYAANVASMTYPLIPNELVSMQAGSAVNLAMPSPGLGGACHAQL 779  
QY 237 NEKHLRSRSISRWVNRNFTPTFRAROKAISTLGASRG 274  
DB 780 SDAGHLS-----WE---QPLKIGQELFMLDLPTPG 806

RESULT 14

US-09-413-814-42  
Sequence 42, Application US/09413814  
Patent No. 6225064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bloecker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hofle, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

CURRENT FILING DATE: 1999-10-07

EARLIER APPLICATION NUMBER: DE 198 46 493.2

EARLIER FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 107

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 42

LENGTH: 2539

TYPE: PRT

ORGANISM: Sorangium cellulosum

Query Match 4.8%; Score 77; DB 4; Length 2539;

Best Local Similarity 22.2%; Pred. No. 78;

Matches 42; Conservative 20; Mismatches 57; Indels 70; Gaps 8;

QY 86 HIYVALKNPVCLTDARRRPINLLARVEQGLCDVGLGDASYGHRITKNPLSTAHATLMGP 145  
DB 1586 HLYOA-----MAAGYSHRIRLLI-VSCGVYDTGAE-----PQPARATMGP 1628  
QY 146 ADAL---YELRALHT-----LDETHALPEAGNRRRVNTRSTVGRNVTL-----F 187  
DB 1629 ALCTPDQYPHLETSHVDLGVVAHDELHAAROLDLSLAECISATVAERQALRGRRHMLDY 1688





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:33:40 ; Search time 20.69 Seconds  
(without alignments)  
100.347 Million cell updates/sec

Title: US-09-720-583a-3

Perfect score: 420  
Sequence: 1 MTTRERLPRNGYSIAAAAK.....EVGVSVGVYVNLNKRDTA 85

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67.5	16.1	375	2 US-08-446-875-8	Sequence 8, Appli
2	67.5	16.1	375	2 US-08-102-385G-8	Sequence 8, Appli
3	62	14.8	328	2 US-08-956-012-1	Sequence 1, Appli
4	61.5	14.6	130	2 US-08-494-907-20	Sequence 20, Appli
5	61.5	14.6	130	2 PCT-US96-10986-20	Sequence 20, Appli
6	61.5	14.6	202	2 US-08-494-907-16	Sequence 16, Appli
7	61.5	14.6	202	5 PCT-US96-10986-16	Sequence 16, Appli
8	61	14.5	100	2 US-08-160-524A-12	Sequence 12, Appli
9	60.5	14.4	114	2 US-08-473-020A-26	Sequence 26, Appli
10	60	14.3	872	1 US-08-491-357-3	Sequence 3, Appli
11	60	14.3	872	3 US-08-968-633-3	Sequence 3, Appli
12	60	14.3	872	3 US-09-196-466-3	Sequence 3, Appli
13	60	14.3	872	5 PCT-US96-10823-3	Sequence 3, Appli
14	59	14.0	450	3 US-09-071-444-1	Sequence 1, Appli
15	58	13.8	977	3 US-08-335-844A-22	Sequence 22, Appli
16	57.5	13.7	553	4 US-09-413-814-3	Sequence 3, Appli
17	57	13.6	261	1 US-08-245-688-2	Sequence 2, Appli
18	57	13.6	261	1 US-08-245-688-6	Sequence 6, Appli
19	57	13.6	261	1 US-08-245-688-12	Sequence 12, Appli
20	56.5	13.5	261	1 US-08-245-688-4	Sequence 4, Appli
21	56.5	13.5	261	1 US-08-245-688-8	Sequence 8, Appli
22	56.5	13.5	261	1 US-08-245-688-10	Sequence 10, Appli
23	56.5	13.5	288	1 US-08-375-709-9	Sequence 9, Appli
24	56.5	13.5	288	1 US-08-752-929-9	Sequence 9, Appli
25	56.5	13.5	288	4 US-09-090-793-6	Sequence 6, Appli
26	56.5	13.5	350	4 US-09-655-270A-17	Sequence 17, Appli
27	56.5	13.5	350	4 US-09-651-941-21	Sequence 21, Appli

28	56.5	13.5	781	2 US-08-675-631-3	Sequence 3, Appli
29	56	13.3	80	4 US-08-737-226-5	Sequence 5, Appli
30	56	13.3	245	4 US-09-321-276-2	Sequence 2, Appli
31	56	13.3	245	4 US-08-916-481-5	Sequence 5, Appli
32	56	13.3	410	3 US-08-948-997-2	Sequence 2, Appli
33	56	13.3	410	4 US-09-348-817A-2	Sequence 2, Appli
34	55.5	13.2	444	4 US-08-822-774-52	Sequence 52, Appli
35	55.5	13.2	444	4 US-09-632-711-52	Sequence 52, Appli
36	55	13.1	244	2 US-08-553-633A-5	Sequence 3, Appli
37	55	13.1	298	2 US-09-006-535-3	Sequence 3, Appli
38	55	13.1	1201	4 US-09-098-901-2	Sequence 2, Appli
39	54.5	13.0	573	1 US-08-200-512-2	Sequence 2, Appli
40	54.5	13.0	2756	1 US-08-375-709-11	Sequence 11, Appli
41	54.5	13.0	2756	1 US-08-752-929-11	Sequence 11, Appli
42	54.5	13.0	2756	4 US-09-090-793-7	Sequence 7, Appli
43	53.5	12.7	224	1 US-08-608-241-4	Sequence 4, Appli
44	53.5	12.7	224	1 US-08-922-182-4	Sequence 4, Appli
45	53.5	12.7	224	2 US-08-919-953-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-446-875-8  
Sequence 8, Application US/08446875  
Patent No. 5858751  
GENERAL INFORMATION:  
APPLICANT: Paulson, James C.  
APPLICANT: Wen, Xiaohong  
APPLICANT: Livingston, Brian Duane  
APPLICANT: Gillespie, William  
APPLICANT: Kelm, Sorge  
APPLICANT: Burlingame, Alma L.  
APPLICANT: Medzhradszky, Katalin  
TITLE OF INVENTION: Compositions and Methods for the  
NUMBER OF SEQUENCES: 16  
TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Poms, Smith, Lande & Rose  
STREET: 2029 Century Park East, Suite 3800  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,875  
FILING DATE: July 12, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/102,385  
FILING DATE: August 4, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oldenkamp, David J.  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 111-197  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-446-875-8

Query Match 16.1%; Score 67.5; DB 2; Length 375;  
Best Local Similarity 25.3%; Pred. No. 0.75;  
Matches 24; Conservative 13; Mismatches 31; Indels 27; Gaps 4;

QY 2 TTRERLRNGYSIAAAKLGVS-----ESTVWRMSEPREEVAVARHARIRLRSR 56  
DB 52 SNRAEVINGSSLPAAVANSNESLKHSTDPASSKW-----RHNOTLSLRIR 97

QY 57 GOSMRAIAAEVGSV--GT-----VHYALNKNRT 83  
DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNRRNST 132

RESULT 2  
US-08-102-385G-8

; Sequence 8, Application US/08102385G  
; Patent No. 5962294

; GENERAL INFORMATION:

; APPLICANT: Paulson, James C.

; APPLICANT: Wen, Xiaohong

; APPLICANT: Livingston, Brian Duane

; APPLICANT: Gillespie, William

; APPLICANT: Kelm, Soreg

; APPLICANT: Burlingame, Alma L.

; TITLE OF INVENTION: Compositions and Methods for the

; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases

; NUMBER OF SEQUENCES: 36

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Openheimer Wolff & Donnelly LLP

; STREET: 2029 Century Park East, 38th Floor

; CITY: Los Angeles

; STATE: CA

; COUNTRY: USA

; ZIP: 90067

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/102,385G

; FILING DATE: 04-AUG-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/925369

; FILING DATE: 04-AUG-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Oidenkamp, David J.

; REGISTRATION NUMBER: 29,421

; REFERENCE/DOCKET NUMBER: 97-062

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (310)788-5000

; TELEFAX: (310) 277-1297

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 375 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-102-385G-8

Query Match 16.1%; Score 67.5; DB 2; Length 375;  
Best Local Similarity 25.3%; Pred. No. 0.75;

Matches 24; Conservative 13; Mismatches 31; Indels 27; Gaps 4;

QY 2 TTRERLRNGYSIAAAKLGVS-----ESTVWRMSEPREEVAVARHARIRLRSR 56  
DB 52 SNRAEVINGSSLPAAVANSNESLKHSTDPASSKW-----RHNOTLSLRIR 97

QY 57 GOSMRAIAAEVGSV--GT-----VHYALNKNRT 83  
DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNRRNST 132

DB 98 KQIKFLNAEKINISVLKGTLPKGNITHYIFNRRNST 132

RESULT 3  
US-08-956-012-1

; Sequence 1, Application US/08956012  
; Patent No. 5972660

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer

; APPLICANT: Lal, Preeti

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: HUMAN HYDROXYPIRUVATE REDUCTASE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: IBM Compatible

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/956,012

; FILING DATE: Filed Herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0410 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 328 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-956-012-1

Query Match 14.8%; Score 62; DB 2; Length 328;  
Best Local Similarity 26.4%; Pred. No. 3.3;  
Matches 23; Conservative 16; Mismatches 28; Indels 20; Gaps 4;

QY 6 RLPRNGYSIAAAKKGVSSTVWRMSE---PREEVAVARHARIRI-----REL 53  
DB 14 RPAEGRVALARA-----ADCEVEQWSDDEPIPAKELRGVAGAGHGLCLSDHVDKRL 68

QY 54 RSEQSMRAIAAEVGSVGTVHYALNK 80  
DB 69 DAAGANLKVIST---MSVGIDHLALDE 92

RESULT 4  
US-08-494-907-20

; Sequence 20, Application US/08494907  
; Patent No. 5955298

; GENERAL INFORMATION:

; APPLICANT: Thomasow, Linda S

; APPLICANT: Bangera, Mahalaxmi

; APPLICANT: Weller, David M

; APPLICANT: Cook, R. James

; TITLE OF INVENTION: Sequences for production of

; TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods

; NUMBER OF SEQUENCES: 20





DB 36 SGLSTIEAVARRAGASKPTIYRMGNK-----AALLIEVESESEQIRKEPKSGSFE 87  
OY 57 -----GOSMRAIAEYGVSVGV 74  
DB 88 NINFLILNIMKVMRETICGEAFRCVIAEAOIDPSTL 123

RESULT 7  
PCT-US96-10986-16  
Sequence 16, Application PC/TUS9610986  
GENERAL INFORMATION:  
TITLE OF INVENTION: Sequences for Production of  
TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stephan A. Pendorf, DOMINIK & STEIN  
STREET: 600 N. West Shore Boulevard, Suite 1000  
CITY: Tampa  
STATE: FL  
COUNTRY: USA  
ZIP: 33609  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10986  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pendorf, Stephan A.  
REGISTRATION NUMBER: 32665  
REFERENCE/DOCKET NUMBER: A700.320  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (813) 289-2966  
TELEFAX: (813) 289-2967  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 202 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-10986-16

Query Match 14.6%; Score 61.5; DB 5; Length 202;  
Best Local Similarity 21.9%; Pred. No. 2;  
Matches 21; Conservative 11; Mismatches 25; Indels 39; Gaps 2;

OY 10 NGYSIAAAKKGIVSESTYKRWTS-----EPREFFVARVAARHARIRELNSEGS 56  
DB 36 SGLSTIEAVARRAGASKPTIYRMGNK-----AALLIEVESESEQIRKEPKSGSFE 87  
OY 57 -----GOSMRAIAEYGVSVGV 74  
DB 88 NINFLILNIMKVMRETICGEAFRCVIAEAOIDPSTL 123

RESULT 8  
US-08-160-524A-12  
Sequence 12, Application US/08160524A  
Patent No. 5851761  
GENERAL INFORMATION:  
APPLICANT: McAdam, Ruth Anne  
APPLICANT: Dale, Jeremy W.  
APPLICANT: Zainuddin, Zainul Fadziruddin B.  
APPLICANT: Catly, David  
TITLE OF INVENTION: PROBES, KITS AND METHODS FOR THE  
TITLE OF INVENTION: DETECTION AND DIFFERENTIATION OF MYCOBACTERIA  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert,  
ADDRESSEE: Attn: Walter H. Dreger  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/160,524A  
FILING DATE: 01-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/752,661  
FILING DATE: 18-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8903968.9  
FILING DATE: 22-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 900411.0  
FILING DATE: 09-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CB90/00276  
FILING DATE: 22-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-55387-1/WHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
US-08-160-524A-12

Query Match 14.5%; Score 61; DB 2; Length 100;  
Best Local Similarity 23.1%; Pred. No. 0.88;  
Matches 18; Conservative 17; Mismatches 31; Indels 12; Gaps 2;

OY 7 LPRNGYSIAAAKKGIVSESTYKRWTS-----EPREFFVARVAARHARIRELNSEGS 59  
DB 20 VEPNQISISGASRLSLPGTIGQWVTAAARKGLGTPGSRTVALESEIQLRKALNEARL 79  
OY 60 MRAIAEYGVSVGVHXA 77  
DB 80 ERDI-----LKRATAYFA 92

RESULT 9  
US-08-473-020A-26  
Sequence 26, Application US/08473020A  
Patent No. 5877273  
GENERAL INFORMATION:  
APPLICANT: Hance, Allan J  
APPLICANT: Grandchamp-Desraux, Bernard  
APPLICANT: Levy-Frebault, Veronique  
APPLICANT: Gicquel, Brigitte  
TITLE OF INVENTION: Nucleotide sequences of actinomycetales,  
TITLE OF INVENTION: applications to the synthesis or detection of nucleic  
TITLE OF INVENTION: acids, products of the expression of such sequences and  
TITLE OF INVENTION: application as immunogenic compositions.  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Walter H. Dreger



HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-968-633-3

Query Match 14.3%; Score 60; DB 3; Length 872;  
Best Local Similarity 34.5%; Pred. No. 23;  
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVAARHARIRIELRSEGSOMRAIAEYGVGVYHALNK 80  
| | | | | : | | | | : | | | | : | | | | : | | | | :  
DB 493 RSTSEQPEPPVODLKAAVAAGVAVHLELFARS--AVSSATHTSDRTIHLAKLSR 545

RESULT 12  
US-09-196-466-3  
Sequence 3, Application US/09196466  
Patent No. 6124434  
GENERAL INFORMATION:

APPLICANT: Golemis, Erica A.  
APPLICANT: Law, Susan  
APPLICANT: Estojak, Joanne  
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL  
TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL  
TITLE OF INVENTION: ALTERATIONS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA

COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196.466  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/491,357  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-196-466-3

Query Match 14.3%; Score 60; DB 3; Length 872;  
Best Local Similarity 34.5%; Pred. No. 23;  
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVAARHARIRIELRSEGSOMRAIAEYGVGVYHALNK 80  
| | | | | : | | | | : | | | | : | | | | : | | | | :  
DB 493 RSTSEQPEPPVODLKAAVAAGVAVHLELFARS--AVSSATHTSDRTIHLAKLSR 545

RESULT 13

PCT-US96-10823-3  
Sequence 3, Application PC/TUS9610823  
GENERAL INFORMATION:

APPLICANT: Golemis, Erica A.  
APPLICANT: Law, Susan  
APPLICANT: Estojak, Joanne  
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL  
TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL  
TITLE OF INVENTION: ALTERATIONS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA

COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10823  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US96-10823-3

Query Match 14.3%; Score 60; DB 5; Length 872;  
Best Local Similarity 34.5%; Pred. No. 23;  
Matches 19; Conservative 10; Mismatches 20; Indels 6; Gaps 2;

QY 30 RWTSEPRE---EFVARVAARHARIRIELRSEGSOMRAIAEYGVGVYHALNK 80  
| | | | | : | | | | : | | | | : | | | | : | | | | :  
DB 493 RSTSEQPEPPVODLKAAVAAGVAVHLELFARS--AVSSATHTSDRTIHLAKLSR 545

RESULT 14  
US-09-071-434-1  
Sequence 1, Application US/09071434  
Patent No. 6080848  
GENERAL INFORMATION:

APPLICANT: Sven, H  
APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN BRAIN-ASSOCIATED PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,434  
FILING DATE: Filed Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrione, Michael C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0517 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAINOT03  
CLONE: 662575  
US-09-071-434-1

Query Match 14.0%; Score 59; DB 3; Length 450;  
Best Local Similarity 31.9%; Pred. No. 13;  
Matches 22; Conservative 11; Mismatches 30; Indels 6; Gaps 3;

QY 14 IAAAKKLG---SESVKRMTSRPREFVARVARHARRIRELSEGSMAITAEVGS 70  
DB 140 VLETVRLGSAVTELTLEEVLEPRTELVA--AARGAR-RQAEAAQLOGLAFWOGVP 196  
QY 71 VGTIVYVALN 79  
DB 197 LSPLOVAEN 205

RESULT 15  
US-08-335-844A-22  
Sequence 22; Application US/08335844A  
Patent No. 6066503  
GENERAL INFORMATION:  
APPLICANT: GRAHAM, MARGARET  
APPLICANT: SMITH, TREVOR STANLEY  
APPLICANT: MUNN, EDWARD ALBERT  
APPLICANT: KNOX, DAVID PATRICK  
APPLICANT: OLIVER, JOANNA JANE  
APPLICANT: NEWTON, SUSAN ELIZABETH  
TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF  
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
STREET: Suite 701-E, 555 Thirteenth St., N.W  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/335,844A  
FILING DATE: 09-JAN-1995

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB PCT/GB93/00943  
FILING DATE: 06-MAY-1993  
PRIOR APPLICATION DATA: GB 9209936  
APPLICATION NUMBER: GB 9209936  
FILING DATE: 08-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, Barbara W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 1181-223A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 977 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-335-844A-22

Query Match 13.8%; Score 58; DB 3; Length 977;  
Best Local Similarity 28.1%; Pred. No. 50;  
Matches 16; Conservative 14; Mismatches 19; Indels 8; Gaps 2;

QY 24 SESVKKRMTSRPREFVARVARHARRIRELSEGSMAITAEVGSVGTIVYVALN 80  
DB 3 AESQEOETQOPRKNYVLRLLTP---IKSL---FALLVAAAVGLSIGLYTFTRK 51

Search completed: October 4, 2002, 15:33:41  
Job time: 33 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:36:07 ; Search time 46.38 Seconds

(without alignments)  
317.045 Million cell updates/sec

Title: US-09-720-583A-3

Perfect score: 420  
Sequence: 1 MTRERLRPRNGYSIAAAKK.....EVGVSVGVTHYALNKNRTDA 85

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPRREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	420	100.0	85	2	09F6K0
2	215	51.2	87	2	046076
3	162	38.6	67	2	09F4V5
4	140	33.3	93	2	09LBV5
5	121	28.8	94	2	09F5N5
6	110	26.2	93	2	030682
7	96	22.9	114	2	09L456
8	82.5	19.6	105	2	045287
9	82	19.5	119	2	052144
10	79	18.8	87	2	093Q22
11	78	18.6	132	2	09EX03
12	78	18.6	150	17	097ZM9
13	78	18.6	318	2	086758
14	78	18.6	318	2	09RJU2
15	78	18.6	338	2	09X467
16	78	18.6	339	2	09K393

17	78	18.6	339	2	09L0Y8	0910Y8 streptomyce
18	78	18.6	339	2	09L0E4	0910E4 streptomyce
19	78	18.6	941	2	09Z1W5	09Z1W5 streptomyce
20	76	18.1	328	2	P94642	P94642 corynebacte
21	76	18.1	329	2	030997	030997 corynebacte
22	75.5	18.0	314	2	085929	085929 sphingomona
23	75	17.9	293	16	09ZRW5	09ZRW5 rhizobium m
24	74.5	17.7	195	2	09ZHM7	09ZHM7 brucella me
25	74.5	17.7	674	2	047763	047763 enterococcu
26	74.5	17.7	679	2	09S6C2	09S6C2 enterococcu
27	74	17.6	139	17	09Y132	09Y132 pyrococcus
28	74	17.6	310	16	09Z1V4	09Z1V4 rhizobium m
29	73.5	17.5	128	2	085930	085930 sphingomona
30	73	17.4	102	16	098B11	098B11 rhizobium 1
31	73	17.4	122	2	088091	088091 escherichia
32	72.5	17.3	184	16	09ZS36	09ZS36 rhizobium m
33	72.5	17.3	315	2	0989L5	0989L5 rhizobium m
34	72.5	17.3	315	2	09X8B3	09X8B3 rhizobium m
35	72.5	17.3	643	17	090Y87	090Y87 pyrococcus
36	72.5	17.3	673	2	047751	047751 enterococcu
37	72.5	17.3	673	2	047783	047783 enterococcu
38	72.5	17.3	678	2	059438	059438 enterococcu
39	72.5	17.3	678	2	047759	047759 enterococcu
40	72.5	17.3	678	2	047801	047801 enterococcu
41	72.5	17.3	678	2	093B65	093B65 enterococcu
42	72.5	17.3	678	2	093BP3	093BP3 enterococcu
43	72	17.1	113	2	068709	068709 yersinia pe
44	72	17.1	113	2	053005	053005 escherichia
45	72	17.1	125	16	099Q10	099Q10 caulobacter

## ALIGNMENTS

RESULT	1	PRELIMINARY	PRT	85 AA.
ID	09F6K0			
AC	09F6K0			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)			
DE	REP2.			
OS	Propionibacterium freudenreichii.			
OG	Plasmid p545.			
OC	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Propionibacterineae; Propionibacteriaceae;			
OC	Propionibacterium.			
OX	NCBI_TaxID=1744;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LMG16545;			
RX	MEDLINE=21091935; PubMed=11157209;			
RA	Jore J.P.M., van Luijk N., Luiten R.G.M., van der Werf M.J.,			
RA	Fouwels P.H.;			
RT	"Efficient transformation system for Propionibacterium freudenreichii			
RT	based on a novel vector";			
RL	Appl. Environ. Microbiol. 67:499-503(2001).			
DR	EMBL: AF291751; AAG25295.1; -.			
KW	Plasmid.			
SQ	SEQUENCE 85 AA; 9425 MM; 30FBBA897F0CC6F0 CRC64;			
Query Match	100.0%;	Score 420;	DB 2;	Length 85;
Best Local Similarity	100.0%;	Pred. No. 2.2e-35;		
Matches	85;	Conservative 0;	Mismatches 0;	Indels 0;
QY	1	MTRERLRPRNGYSIAAAKLGVSSTVKKMTSEPRFEVAVAAARHARIRREISGQSM 60		
DB	1	MTRERLRPRNGYSIAAAKLGVSSTVKKMTSEPRFEVAVAAARHARIRREISGQSM 60		
QY	61	RAIAAEVGVSVGVTHYALNKNRTDA 85		
DB	61	RAIAAEVGVSVGVTHYALNKNRTDA 85		

RESULT	2			
046076				
ID	046076	PRELIMINARY;	PRF;	87 AA.
AC	046076:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHEETICAL 9.8 KDA PROTEIN.			
OS	Corynebacterium glutamicum (Brevibacterium flavum).			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Corynebacteriaceae;			
OC	Corynebacterium			
OX	NCBI_Taxid-1718;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-1014;			
RA	Shi S.L., Wang Z.X., Deng Y., Zhu H., He B., Zheng Z.X.;			
RT	"Complete Nucleotide Sequence of a Plasmid pXZ10142 from			
RL	Corynebacterium glutamicum."			
DR	Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.			
EMBL:	X72691; CAAS1239.1; -.			
KM	Hypothetical protein.			
SC	SEQUENCE 87 AA; 9769 MW; 86DA0A01601185747 CRC64;			

	Query Match	51.2%	Score 215	DB 2	Length 87
	Best Local Similarity	53.3%	Pred. No. 1.3e-14		
	Matches	40	Conservative 17	Mismatches 18	Indels 0
				Gaps	0
QY	1	MTTRETLRNPGSGTAAAKKGVSESTYKRWKSPSEPRENVAAARHARIRLRSEGGSM	60		
DB	1	MTKRRIRTRNGTTLIEVADGCTGSLTATLERMTSAPREDITLAQANERKRYAVQDLRANKGLSM	60		
QY	61	RAIAAEVGVSGTVH	75		
DB	61	RAIAAIGCGSVGLVH	75		

RESULT	3			
09FAV5				
ID	09FAV5	PRELIMINARY;	PRT;	67 AA.
AC	09FAV5			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last Sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	REPB.			
GN	REPB.			
OS	Brevibacterium linens.			
OG	Plasmid LIM.			
OC	Actinomycetales; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Micrococciaceae; Brevibacteriaceae; Brevibacterium			
OK	NCBI TaxID=1703;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Moore M.W., Bowling D., Glenn D.;			
RT	"Nucleotide sequence of a plasmid from Brevibacterium linens:			
RT	Construction of a cloning vector."			
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AY004211; AAF89086.1; -.			
KW	Plasmid.			
50	SEQUENCE	67 AA;	7850 MW;	CF7DDE3EA6EA4170 CRC64;

```

Query Match      38.6%   Score 162: DB 2: Length 67:
Best Local Similarity 54.2%:   Pred. No. 2,3e-09:
Matches 32: Conservative 7: Mismatches 20: Indels 0: Gaps 0:

OY 1 MTTTRRLPRNGYSIAAAAKKLCVSESTYKRYMTSEPRREEVAAVAAHARIRLRSGQS 59
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 MTAERHPHPNGSVLELAKVGVSEKSYLRMTSEPREVLLSAAGDRAVRIIRLRREGLS 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT	4	
09LBJ5		
ID	09LBJ5	PRELIMINARY;
AC	09LBJ5	PRT; 93 AA.
DT	01-OCT-2000	(TREMBLrel. 15, Created)
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	DNA-BINDING REPLICATION PROTEIN.	
GN	REPB.	
OS	Rhodococcus rhodochrous.	
OC	Plasmid PRC4.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;	
OC	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.	
OX	NCBI_TaxID:1829;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-IF03338;	
RX	MEDLINE=21196497;	PubMed=11302154;
RA	Hirasawa K., Ishii Y., Kobayashi M., Koizumi K., Maruhashi K.;	
RT	"Improvement of Deulfurization Activity in Rhodococcus erythropolis	
RT	KA2-5-1 by Genetic Engineering."	
RL	Biosci. Biotechnol. Biochem.	65:239-246(2001).
RL	EMBL: AB040101; BAA92819.1;	
DR	InterPro: IPR000792;	HTH_LuxR.
KW	DNA-binding; Plasmid.	
SO	SEQUENCE	93 AA; 10124 MW; D328D70CFA51A21A CRC64;

[illegible]

	RESULT	5
09F5N5	09F5N5	PRELIMINARY; PRT; 94 AA.
ID	09F5N5	
AC	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	PUTATIVE DNA-BINDING PROTEIN.	
OS	Rhodococcus erythropolis.	
OG	Plasmid pN30.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.	
OX	NCBI_TaxID=1833;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ryabchenko L.E., Novikov A.D., Golyshtin P.N., Yansenko A.S.;	
RT	"Rhodococcus erythropolis Plasmid pN30 putative replicase (ORF1), and	
RT	putative DNA-binding replication protein (ORF2).";	
RL	Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.	
DR	EMBL: AF312210; AAC29855.1; "	
KW	DNA-binding; Plasmid.	
SQ	SEQUENCE 94 AA; 10409 MW; B3452537C39166E7 CRC64;	

Query Match	28.8%	Score 121	DB 2	Length 94			
Best Local Similarity	37.8%	Pred. No. 4.9e-05					
Matches	28	Conservative 14	Mismatches 33	Indels 0			
			Gaps 0				
Qy	9	RNCGYIAAAAKKLGVS	ESTKRWTS	EPREEFYAVYAAARHARIRLR	SSGQSHRAITAEV	G 68	
Db	9	RKKMTAAAAAEFFGAST	RIORLFAEP	RNDYDYGRAK	KARDKVALR	KGKLYREIAE	ME 68



```

OY 69 VSVGTVHYALNKNR 82
      :| | | | :| |
Db 69 ISTGIVGRLLHDAR 82

RESULT 6
030682 ID 030682 PRELIMINARY: PRT: 93 AA.
030682 AC 030682:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE DNA-BINDING REPLICATION PROTEIN.
GN REPB.
OS Rhodococcus erythropolis.
OC Plasmid PFAJ2600.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NI86/21;
RA MEDLINE=98015402; PubMed=9353918;
RA De Mot R., Nagy I., De Schrijver A., Pattanapitpaisal P.,
RA Schoofs G., Vanderleyden J.;
RT "Structural analysis of the 6-kb cryptic plasmid PFAJ2600 from
RT Rhodococcus erythropolis NI86/21 and construction of Escherichia coli-
RT Rhodococcus shuttle vectors."
RL Microbiology 143:3137-3147(1997).
RW EMBL; AF015088; AAC45811.1; -.
KW DNA-binding; Plasmid.
SQ SEQUENCE 93 AA: 10672 MW; 74FA8C05CE8E5ED CRC64;

Query Match 26.2%; Score 110; DB 2; Length 93;
Best Local Similarity 35.7%; Pred. No. 0.00063;
Matches 25; Conservative 12; Mismatches 33; Indels 0; Gaps 0;

OY 5 ERLPRNGSIAAAKRLGSESTVSRKWTSEPEEEYARVAANHARKRELRSQGSRAIA 64
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 ENQTRRNRTAKEMARLGVSEPTIRNIVAESDSYQARAERDRPAVKLRQGMRYREIA 63
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 65 AEVGSVGTV 74
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 EEMEISTGAV 73

RESULT 7
091456 ID 091456 PRELIMINARY: PRT: 114 AA.
091456 AC 091456:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHEICAL 13.0 KDA PROTEIN (REPLICATION PROTEIN).
GN REPB.
OS Propionibacterium jensenii, and
OS Propionibacterium acidipropionict.
OC Plasmid pLME106, and Plasmid pRG01.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Propionibacterineae; Propionibacteriaceae;
OC Propionibacterium.
OX NCBI_TaxID=1749, 1748;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=P. jensenii; STRAIN=Df1; PLASMID=pLME106;
RA Stierli M.P., Meile L., Teuber M.;
RT "Molecular analysis of plasmid pLME106, a plasmid isolated from
RT Propionibacterium jensenii."
RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```

Query Match	22.9%	Score 96	DB 2	Length 114
Best Local Similarity	29.3%	Pred. No. 0.021		
Matches 24	Conservative 17	Mismatches 39	Indels 2	Gaps 2
Qy	4	RRRLPRNGYSIAAAKKGVSSTYKRTSREFFVAVARHARIRELR-SEGOSMA	62	
Db	5	QHRVRRRG-TAREAAERGVASIRTAQRTSIPREMIITOKAVEBEIRAYKYDEGHTWGE	63	
Qy	63	IAAEVGSVGTVHVVALNKRTD	84	
Db	64	TSRHFGIAKTTAQERARRARE	85	
RESULT	8			
ID	045287	PRELIMINARY:	PRT:	105 AA.
AC	045287			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	ORF1 PROTEIN.			
GN	ORF1.			
OS	Bifidobacterium longum.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.			
OX	NCBI_TaxID=1679;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=B2577;			
RX	MEDLINE=96341130; PubMed=8761732;			
RA	Rossi M., Brigidi P., Gonzalez Varea Y Rodriguez A., Matteuzzi D.;			
RT	"Characterization of the plasmid pMB1 from Bifidobacterium longum and			
RL	its use for shuttle vector construction."			
Res.	Microbiol. 147:133-143(1996).			
EMBL	X84655; CAA59149.1; -			
SO	SEQUENCE 105 AA; 12317 MW; 295DDI5830188041 CRC64;			
Query Match	19.6%	Score 82.5	DB 2	Length 105
Best Local Similarity	28.7%	Pred. No. 0.44		
Matches 25	Conservative 13	Mismatches 42	Indels 7	Gaps 3
Qy	2	TTRRLPRNGYSIAAAKKGVSSTYKRTSREFFVAVARHARIRELR-SEGOSM	60	
Db	5	TIRKRPRPSARLALAA--YGVSTTIGSVAMAKKREDWIDQAAARBAVRSYHDEGHTW	61	
Qy	61	RAIAAEVGSVGTVH--YALNKRTD	84	
Db	62	PQTAEHFNNISGCAVRORCYRARKERED	88	
RESULT	9			
ID	052144	PRELIMINARY:	PRT:	119 AA.
AC	052144			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL 13.7 KDA PROTEIN.			

Query Match	Best Local Similarity	19.5%;	Score 82;	DB 2;	Length 119;
Matches	26;	Conservative 16;	Matches 8;	Mismatches 34;	Indels 4;
Gaps	2;				
OS	Mycobacterium fortuitum.				
OC	Plasmid pAL5000.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1766;				
RP	(1)				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=89138007; PubMed=3224826;				
RA	Rauzier J., Moniz-Pereira J., Gilequel-Sanzey B.;				
RT	"Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium				
RT	fortuitum.";				
RL	Gene 71:315-321(1988).				
DR	EMBL; M23557; AAA98172.1; "				
KW	Hypothetical protein; Plasmid.				
SO	SEQUENCE 119 AA; 13683 MW; 282F8087C71FCAB CRC64;				
Query Match	Best Local Similarity	19.5%;	Score 82;	DB 2;	Length 119;
Matches	26;	Conservative 16;	Matches 8;	Mismatches 34;	Indels 4;
Gaps	2;				
OS	Mycobacterium fortuitum.				
OC	Plasmid pAL5000.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1766;				
RP	(1)				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=89138007; PubMed=3224826;				
RA	Rauzier J., Moniz-Pereira J., Gilequel-Sanzey B.;				
RT	"Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium				
RT	fortuitum.";				
RL	Gene 71:315-321(1988).				
DR	EMBL; M23557; AAA98172.1; "				
KW	Hypothetical protein; Plasmid.				
SO	SEQUENCE 119 AA; 13683 MW; 282F8087C71FCAB CRC64;				
Query Match	Best Local Similarity	18.8%;	Score 79;	DB 2;	Length 87;
Matches	19;	Conservative 35.8%;	Pred. No. 0.81;	Mismatches 21;	Indels 4;
Gaps	1;				
OS	Mycobacterium fortuitum.				
OC	Plasmid pAL5000.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1766;				
RP	(1)				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=89138007; PubMed=3224826;				
RA	Rauzier J., Moniz-Pereira J., Gilequel-Sanzey B.;				
RT	"Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium				
RT	fortuitum.";				
RL	Gene 71:315-321(1988).				
DR	EMBL; M23557; AAA98172.1; "				
KW	Hypothetical protein; Plasmid.				
SO	SEQUENCE 87 AA; 10211 MW; 69505094D53080DE CRC64;				
Query Match	Best Local Similarity	18.8%;	Score 79;	DB 2;	Length 87;
Matches	19;	Conservative 35.8%;	Pred. No. 0.81;	Mismatches 21;	Indels 4;
Gaps	1;				
OS	Mycobacterium fortuitum.				
OC	Plasmid pAL5000.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1766;				
RP	(1)				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=89138007; PubMed=3224826;				
RA	Rauzier J., Moniz-Pereira J., Gilequel-Sanzey B.;				
RT	"Complete nucleotide sequence of pAL5000, a plasmid from Mycobacterium				
RT	fortuitum.";				
RL	Gene 71:315-321(1988).				
DR	EMBL; M23557; AAA98172.1; "				
KW	Hypothetical protein; Plasmid.				
SO	SEQUENCE 87 AA; 10211 MW; 69505094D53080DE CRC64;				

```

RN      [1]          SEQUENCE FROM N.A.
RP      STRAIN-A3(2);
RC      Saunders D.C., Harris D.;
RA      Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RM      [2]
RN      [2]          SEQUENCE FROM N.A.
RP      STRAIN-A3(2);
RC      Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA      Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RM      [3]
RN      [3]          SEQUENCE FROM N.A.
RP      STRAIN-A3(2);
RC      MEDLINE=97000351; PubMed=8643436;
RA      Redenbach M., Kleiser H.M., Denapite D., Eichner A., Cullum J.,
RA      Kinasli H., Hopwood D.A.;
RT      "A set of ordered cosmids and a detailed genetic and physical map for
RL      Mol. Microbiol. 21:77-96(1996).
DR      EMBL; AL445503; CACI3075.1; -.
SQ      Hypothetical protein.
KW      SEQUENCE 132 AA; 14688 MW; 08522DDB64286C3C CRC64;

Query Match              18.6%; Score 78; DB 2; Length 132;
Best Local Similarity    41.5%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 25; Indels 4; Gaps 3;

OY      15 AAAAKKILGV-SESTVKKRWT--EPREEFVARVAARHARIRELRSEGQSMR-AIAAEVGVS 70
        | ||||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      13 ATAAKKILGVLEATPAEREGCVTTAEALNALQADPPAIRLRNRGDPHPRPVAAKLGV 72

OY      71 VGVVH 75
        : : |
DB      73 IAGLH 77

RESULT  12
Q97ZW9 PRELIMINARY; PRT; 150 AA.
ID      Q97ZW9
AC      Q97ZW9
DT      01-OCT-2001 (TReMbUrel. 18, Created)
DT      01-OCT-2001 (TReMbUrel. 18, Last sequence update)
DT      01-OCT-2001 (TReMbUrel. 18, Last annotation update)
DE      HYPOTHEMETICAL PROTEIN SS00454.
GN      SS00454.
OS      Sulfolobus solfataricus.
OC      Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX      NCBI_TaxID=2287;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 35092 / DSM 1617 / P2;
RX      MEDLINE=21332296; PubMed=11427726;
RA      She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA      Ameyor M.J., Chan-Weiner C.C.Y., Clausen I.G., Curtis B.A.,
RA      De Moors A., Eruso G., Fletcher C., Gordon P.M.K.,
RA      Heikamp-de Jong I., Jeffries A.C., Kozeza C.J., Medina N., Peng X.,
RA      Tra-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA      Charibolis R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA      Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT      "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR      EMBL; AE006677; AAK40779.1; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 150 AA; 17577 MW; E14E9250440E56B6 CRC64;

Query Match              18.6%; Score 78; DB 17; Length 150;
Best Local Similarity    28.6%; Pred. No. 1.9;
Matches 24; Conservative 17; Mismatches 37; Indels 6; Gaps 2;

OY      5 ERLPRNGYSIAAAKKILGVSESTVKRWTSSEPREEFVARVA-ARRARIRELRSEGQSMRI 63
        :: | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```
Db 11 KMYEGSLRELANOLGSLYSKVRRLIKAKVNFQKVPNDKIÖKIEMKOGYSANRI 70
OY 64 AAEVGSVGTV-----HYALNKNR 82
       |::||
Db 71 SRELINENTVLRILKKYNLGGKR 94

RESULT 13
ID 086758 PRELIMINARY; PRT; 318 AA.
AC 086758;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE TRANSPOSASE.
GN SC6A9.26 OR SCJ11.20.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL109949; CAB52905.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; Rve; 1.
SO SEQUENCE 318 AA; 36530 MW; 61446B9CA1DBEE9 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 318;
Best Local Similarity 27.4%; Pred. No. 4.4;
Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

OY 2 TTRELRPR-----NGYSIAAAKRLGVSESTVKRW-----SEPR-- 36
       | | | | |:::| | | | |
Db 11 TGRRLARCVEDGWPVRAAERFOVSHTTASWARRYROLGYTGMSDRSSRPHTQPRRT 70
OY 37 -----EEFVAARVAARHARIRELRSGOSMRAIAAEVGSVGTVHAL 78
       | | | | |:::| | | | |
Db 71 AAVEEHVLRRLREH-RIGPLR-----LAVRCGIAASTAHRIL 107

RESULT 14
OY 14 PRELIMINARY; PRT; 318 AA.
ID 09RJU2
AC 09RJU2;
```

```
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE TRANSPOSASE (PUTATIVE INSERTION ELEMENT TRANSPOSASE).
GN SCF41.27 OR SC10B8A.13.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117387; CAB55730.1; -.
DR EMBL; AL391454; CAB04115.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; Rve; 1.
SO SEQUENCE 318 AA; 36544 MW; 65546B9CA1DBEE9 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 318;
Best Local Similarity 27.4%; Pred. No. 4.4;
Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

OY 2 TTRELRPR-----NGYSIAAAKRLGVSESTVKRW-----SEPR-- 36
       | | | | |:::| | | | |
Db 11 TGRRLARCVEDGWPVRAAERFOVSHTTASWARRYROLGYTGMSDRSSRPHTQPRRT 70
OY 37 -----EEFVAARVAARHARIRELRSGOSMRAIAAEVGSVGTVHAL 78
       | | | | |:::| | | | |
Db 71 AAVEEHVLRRLREH-RIGPLR-----LAVRCGIAASTAHRIL 107

RESULT 15
OY 15 PRELIMINARY; PRT; 338 AA.
ID 09X467
AC 09X467;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE TRANSPOSASE.
GN TNP4.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN-A3(2);  
 RX MEDLINE-99250253; PubMed-10231572;  
 RA Chung H.J., Kim E.J., Suh B., Choi  
 RT "Duplicate genes for Fe-containing superoxide dismutase in  
 ST Streptomyces coelicolor A3(2).";  
 RL Gene 231:87-93(1999);  
 DR EMBL; AF099014; AAD33127.1; -;  
 DR InterPro; IPR01584; Rve.  
 DR Pfam; PF00665; rve; 1.  
 SO SEQUENCE 338 AA; 38807 MW; E92CD48B11D0C615 CRC64;

Query Match 18.6%; Score 78; DB 2; Length 338;  
 Best Local Similarity 27.4%; Pred. No. 4.7;  
 Matches 29; Conservative 11; Mismatches 28; Indels 38; Gaps 5;

QY 2 TTRERLP-NGYSIAAAKKGVSSTVKRMT-----SEPR-- 36  
 Db 32 TGRRLACVYEDGFPVRAAERFVSHHTASRWARRYQOLGVTGMSDRSSRPHHQPRT 91  
 QY 37 ----EEFVARYAARHARIRELRSEGSMAIAAEVGVSGTVHYAL 78  
 Db 92 AAAYEHVLRRLREH-RIGPLR-----LAVRCGIAASTAHRTL 128

Search completed: October 4, 2002, 15:36:08  
 Job time: 180 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:34:40 ; Search time 55.08 Seconds  
(without alignments)  
171.410 Million cell updates/sec

Title: US-09-720-583A-3

Perfect score: 420  
Sequence: 1 MTRERLRPRNGYSIAAAAKR.....EVCVSGTVHAIKMKRTDA 85

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq\_032802:\*

- 1: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	420	100.0	85	21	AAV44637
2	82	19.5	314	18	AAW14834
3	79	18.8	314	13	AAK29623
4	79	18.8	314	13	AAK37873
5	79	18.8	314	13	AAK34544
6	76	18.1	314	13	AAK20991
7	69.5	16.5	316	22	AAAB14150
8	67	16.0	98	22	AAAG91658
9	67	16.0	279	22	AAU44478
10	67	16.0	927	22	AAE10133
11	66	15.7	214	22	AAU59374

12	65.5	15.6	132	22	AAU58347
13	65	15.5	168	22	ABB64106
14	64	15.2	162	22	AAU56724
15	64	15.2	743	22	ABG68417
16	64	15.2	1100	22	ABG24905
17	63.5	15.1	98	22	AAAG90746
18	63.5	15.1	98	22	AAAG90753
19	63.5	15.1	186	18	AAW30754
20	63.5	15.1	186	19	AAW50189
21	63.5	15.1	186	19	AAW44288
22	63.5	15.1	186	19	AAW44289
23	63.5	15.1	186	19	AAW44286
24	63.5	15.1	186	19	AAW44287
25	63.5	15.1	211	22	AAU52836
26	63.5	15.1	326	22	AAU34593
27	63.5	15.1	547	22	AAO13897
28	63.5	15.1	573	21	AAV82154
29	63.5	15.1	580	21	AAV82157
30	63.5	15.1	763	21	AAV82155
31	63.5	15.1	953	21	AAV82156
32	63	15.0	219	22	AAU63583
33	63	15.0	247	21	AAAG20359
34	63	15.0	279	21	AAAG20358
35	63	15.0	284	21	AAAG20357
36	63	15.0	2169	22	AAAG91701
37	62.5	14.9	507	22	AAAG91701
38	62	14.8	158	21	AAV81764
39	62	14.8	186	21	AAAG40791
40	62	14.8	204	21	AAAG40790
41	62	14.8	227	21	AAAG40789
42	62	14.8	328	20	AAV43622
43	62	14.8	328	21	AAV43622
44	62	14.8	356	22	AAU23556
45	62	14.8	362	22	AAU22915

#### ALIGNMENTS

RESULT 1	AAV44637	standard; Protein; 85 AA.
XX	AAV44637:	
XX	07-APR-2000	(first entry)
XX	Proionibacterium LMG 16545	protein-2.
XX	Proionibacterium LMG 16545; vector; plasmid; antigen; vaccine; enzyme;	
KW	nutritional factor; growth factor; clotting factor; antimicrobial; drug;	
KW	hormone; vitamin B12; animal feed; lactic acid bacteria; foodstuff;	
XX	cheese.	
OS	Proionibacterium freudenreichii LMG 16545.	
XX	WO9967356-A2.	
XX	29-DEC-1999.	
PD	25-JUN-1999;	99MO-EP04416.
PF	25-JUN-1998;	98EP-0305033.
XX	(KONN ) GIST-BROCADES BV.	
PA	Pouwels PH, Van Luijk N, Jore JPM, Luiten RGM;	
XX	WPI; 2000-13697/12.	
DR	N-PSDB; AA49691.	
XX	Novel vectors containing Proionibacterium sequences, used to express	
PT	homologous or heterologous proteins	

Claim 16; Page 26; 54pp: English.

The present amino acid sequence is encoded by the endogenous plasmid Proplionibacterium LMG 16545 ORF-2, derived from Proplionibacterium firendonreichi deposited under CBS 101022 or CBS 101023. The vector has insertion sites for foreign DNA fragments and is capable of autonomous replication. It can be used safely as they remain extrachromosomal and are very stable. They can be used to express homologous or heterologous proteins, which may include antigens for use in vaccines, nutritional, growth and clotting factors, antimicrobials, vitamins (especially vitamin B12), enzymes, hormones and drugs. The Proplionibacterium host cells are used in the production of animal feeds, manufacture of foodstuffs like, sausages and cheese, where they may be used instead of or in addition to lactic acid bacteria and in the manufacture of vitamin B12. The DNA sequence can also be used as a source of primers and probes.

Sequence 85 AA:

Query Match 100.0%; Score 420; DB 21; Length 85;  
Best Local Similarity 100.0%; Pred. No. 3.2e-46;  
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 1 MTTTRRLPRNGSIAAAKKILGVSESTVWRKTSSEPRFVAVAAARHARIRLRSGGSM 60  
Db 1 mlttrrlprngsisaakklgvsestvwrtwscprfearvaarharirrlrsgsgm 60  
61 RAIAAEVGVSGTVHVALNKNRTDA 85  
61 raiiaevgvsgtvhvalnknrtda 85

Db 61 raiiaevgvsgtvhvalnknrtda 85

RESULT 2  
AAW14834  
AAW14834 standard; Protein: 314 AA.

AAW14834:  
28-MAY-1997 (first entry)  
Protein encoded by M.rep ORF1 of plasmid pmw101.  
Plasmid pmw101; Mycobacteria; Lipoprotein secretion signal sequence;  
antigen; Bacterial vaccine; Lyme disease; Borrelia burgdorferi;  
surface protein.  
Synthetic.  
US5583038-A.  
10-DEC-1996.  
21-OCT-1991; 91US-0780261.  
17-NOV-1992; 92US-0977630.  
21-OCT-1991; 91US-0780261.  
(MEDI-) MEDIMMUNE INC.  
Stover CK.  
WI: 1997-042315/04.  
N-PSDB; AAT64413.  
Mycobacteria expressing secretion signal of lipoprotein and  
heterologous antigen, esp. outer surface protein A or B of Borrelia  
burgdorferi - are used in the form of a live bacterial vaccines  
against Lyme disease  
Example 1; Fig 5; 122pp: English.  
The sequences given in AAW14834-36 are proteins which are encoded by

```
CC the plasmid pMW101. This plasmid was used in the construction of a
CC series of vectors used to produce the recombinant Mycobacteria of
CC the invention. The recombinant mycobacteria of the invention are
CC transformed with DNA encoding a polypeptide which comprises a
CC lipoprotein secretion signal sequence and an antigen (Ag) heterologous
CC to the mycobacteria. The lipoprotein secretion signal causes the Ag
CC to be produced as a lipoprotein. The mycobacteria may be used in the
CC form of a live bacterial vaccines against Lyme disease, where the
CC bacteria express a surface protein of Borrelia burgdorferi, the
CC causative agent of Lyme disease. This sequence is printed in the
CC specification in the C-terminal to N-terminal orientation as
CC determined from the corresponding DNA sequence.
CC
SQ Sequence 314 AA:

Query Match 19.5%; Score 82; DB 18; Length 314;
Best Local Similarity 36.1%; Pred. No. 0.04;
Matches 26; Conservative 8; Mismatches 34; Indels 4; Gaps 2;

OY 15 AAAAKKGVSESTVKRMTSEPREFVARVAAHARIREL-SEGOSMRAIAEVCVSQGT 73
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 223 aearitgttserhvrylvaqersewlaeqaarrrirayhdgfnswptakfhghldt 282
OY 74 VH---YALNKNR 82
   | | | | |
Db 283 vKrlgyrarker 294

RESULT 3
AAR29623
ID AAR29623 standard; Protein; 314 AA.
XX
AC AAR29623;
XX
DT 04-JUN-1993 (first entry)
XX
DE Tet e gene product from pMW101 - a deletion mutant of pYB125.
XX
KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
KM Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
XX pertussis; malaria; influenza virus; CTL; herpes virus.
OS Mycobacterium.
PN WO9221376-A.
XX
PD 10-DEC-1992.
XX
PF 01-JUN-1992; 92WO-US04538.
XX
PR 06-JUN-1991; 91US-0711643.
PA (MED1-) MEDIMMUNE INC.
XX
PI Dela Cruz V, Stover CK;
XX
DR WPI, 1992-433380/52.
XX
DR N-PsDB; AAO31727.
XX
PT Method of inducing cytotoxic T-lymphocyte response - esp
PT expression products of transformed Mycobacterium are useful as
PT vaccines against HIV, pertussis, malaria, influenza virus, herpes
PT virus, etc.
XX
PS Example 1; Fig 5; 86pp; English.
XX
CC This sequence is the tet e gene encoded by plasmid pMW101.
XX
SQ Sequence 314 AA;
```







XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PS  
XX  
PS Claim 17; SEQ ID NO: 5412; 246bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
CC  
SQ Sequence 98 AA:  
  
Query Match 16.0%; Score 67; DB 22; Length 98;  
Best Local Similarity 31.6%; Pred. No. 0.76;  
Matches 18; Conservative 6; Mismatches 27; Indels 6; Gaps 1;  
  
QY 11 GYSIAAAAKKLGVSESTVKKW-----TSEPREFPARYAARARIRLRSEGSQNR 61  
Db 23 gtsiqkqkandlgimrmlknwltkysansthngantatalasearlrqlekenallr 79  
  
RESULT 9  
AAU44478  
ID AAU44478 standard; Protein; 279 AA.  
XX  
AC AAU44478;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #5374.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI: 2001-616774/71.  
XX  
DR N-PSDB; AAS59522.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX

PS Example 1; SEQ ID NO 5673; 1069bp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 279 AA:

Query Match 16.0%; Score 67; DB 22; Length 279;  
Best Local Similarity 36.5%; Pred. No. 2.9;  
Matches 23; Conservative 11; Mismatches 25; Indels 4; Gaps 1;  
  
QY 22 GVSSTVKKRWTSEPFYARVAARARIRLRSEGSQNRKAIAAEV---GVSQGVTHYA 77  
Db 10 gmatmvtgtsqigrefvtqlaargddlvivardtermaalkadvarygsvevlaad 69  
  
QY 78 LNK 80  
Db 70 lsr 72  
  
RESULT 10  
AAE10133  
ID AAE10133 standard; Protein; 927 AA.  
XX  
AC AAE10133;  
XX  
DT 29-NOV-2001 (first entry)  
XX  
DE Streptomyces noursei nystatin gene, NysR3.  
XX  
KW polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; transcriptional activator.  
XX  
OS Streptomyces noursei.  
XX  
FH Key Location/Qualifiers  
FT Domain 26..47  
FT /label= LZ  
FT /note= "Leucine zipper motif"  
FT Domain 548..568  
FT /label= TM1  
FT /note= "Transmembrane domain"  
FT Domain 583..610  
FT /label= TM2  
FT /note= "Transmembrane domain"  
FT Domain 684..912  
FT /label= HTH  
FT /note= "LuxR-type helix-turn-helix motif (DNA binding)"  
XX  
PN WO200159126-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 08-FEB-2001; 2001WO-GB00509.

```
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITTENSKAPSLIGE.
PA (SAFE) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DIE/) DIEGLEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVNIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Vala S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR N-PSDB: AAD17184.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 15; Page 177-178; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a transcriptional activator encoding Streptomyces noursei
CC nystatin gene, NysR3.
XX
SQ Sequence 927 AA;
XX
Query Match 16.0%; Score 67; DB 22; Length 927;
Best Local Similarity 29.2%; Pred. No. 14;
Matches 26; Conservative 13; Mismatches 36; Indels 14; Gaps 3;
XX
QY 6 RLPNGYSIAAAKKLGSESTVKRWSEPRERF-----VARVAARARIRELRSE 56
Db 825 RLPpgeaaylreaaglaacgywplae-faelglglsatvpvaltlpserivasyvr 883
QY 57 GOSMRATAAEVGSVGTVHYALNKRRTDA 85
Db 884 gltngalateltgvs-----stravekhltsa 908
XX
RESULT 11
AAU59374
ID AAU59374 standard; Protein: 214 AA.
XX
AC AAU59374;
XX
XX 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #20270.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
```

```
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORT-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
DR N-PSDB: AAS59602.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 20569; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 214 AA;
XX
Query Match 15.7%; Score 66; DB 22; Length 214;
Best Local Similarity 36.4%; Pred. No. 2.8;
Matches 24; Conservative 6; Mismatches 26; Indels 10; Gaps 3;
XX
QY 3 TRELRPR---NGYSIAAAKKLGSESTV-----KRWTSRPREFVARV-AAHNAHIRE 52
Db 25 trstkrpspsnglaslftmaklyvtsypesepkrtwtspssthtaltnhadtklfd 84
QY 53 LRSEGO 58
Db 85 lsnnpq 90
XX
RESULT 12
AAU58347
ID AAU58347 standard; Protein: 132 AA.
XX
AC AAU58347;
XX
XX 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #19243.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
```

```

PD 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham Jr, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX
XX N-PSDB; AAS59590.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX
XX Example 1; SEQ ID NO 19542; 1069pp; English.
XX
XX Sequences AAG39105-AAG56801 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 132 AA;
XX
XX Query Match 15.6%; Score 65.5; DB 22; Length 132;
XX Best Local Similarity 28.7%; Pred. No. 1.7;
XX Matches 31; Conservative 10; Mismatches 34; Indels 33; Gaps 4.
XX
XX 9 RNCYSIAAAAKKLTGSESVYRKWMTSEPREFARVAA----- 45
XX | | | | | | | | | | | | | | | | | | | |
XX Db 17 lpyg--lrrmdagvagflvkdtpedlaqvrvkvhaggraldpalaesllghnplsg 74
XX
XX 46 RHARIRELRSEGSQMRATAEYGVSVGVH-----PALNK-----NRTDA 85
XX | | | | | | | | | | | | | | | | | | | |
XX Db 75 rerellrrlaesgasissllaselclsvgytrmrvhsaigtgtaantea 122
XX
XX RESULT 13
XX ID ABB64106
XX ABB64106 standard; Protein; 868 AA.
XX
XX ABB64106;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 19110.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX

```

```

XX Drosophila melanogaster.
PN WO200171042-A2.
PD XX
PD 27-SEP-2001.
PE XX
PF 23-MAR-2001; 2001WO-US09231.
PX XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PA (PEKE ) PE CORP NY.
PI Venter JC, Adams M, Li FWD, Myers EW;
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08209.
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 19110; 21pp + Sequence Listing; English.
PX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CX SQ
SQ Sequence 868 AA:

Query Match 15.5%; Score 65; DB 22; Length 868;
Best Local Similarity 27.9%; Pred. NO. 23;
Matches 19; Conservative 15; Mismatches 26; Indels 8; Gaps 2;

QY 14 IAAAKKLGVESTVYKRWTSSEPREFVARVAARHARIERLRSQGSMRAIAA--EVGVSV 71
   ||| : | : :: | || | : |||| | : || | : ::
Db 746 javahtnqslqkvlrkas-----vagavafggrlyelrtsagheiadcaaqgeifnyi 799
   QY 72 GTVHYALN 79
      | :: |
      Db 800 grylqhnan 807

RESULT 14
ID AAU56724 standard; Protein: 162 AA.
AC AAU56724;
XX AAU56724;
DE 27-FEB-2002 (first entry)
XX Propionibacterium acnes immunogenic protein #17620.
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitits; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked Immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
OS Propionibacterium acnes.
XX WO200181581-A2.
XX 01-NOV-2001.

```

```

PF 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
DR MPI: 2001-616774/71.
XX N-PSDB: AAS59578.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1: SEQ ID NO 17919; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 162 AA:
SQ

```

Query Match 15.2%; Score 64; DB 22; Length 162;  
Best Local Similarity 30.4%; Pred. No. 3.5;  
Matches 17; Conservative 9; Mismatches 14; Indels 16; Gaps 2;

```

QY 12 YSIATAAKKLGSESTVKKWTSEPRFEFVARVAARHARIRE-----LRSEGSQMR 62
DB 31 hrvsdvaallgyvddtvttrwdq-----fhvraakdtgtlrvdgaslaa 75

```

RESULT 15  
ABB68417  
ID ABB68417 standard; Protein: 743 AA.  
XX  
XX ABB68417;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 32043.  
DE  
XX  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF

```

XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI MPI: 2001-656860/75.
XX N-PSDB: ABL12520.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 32043; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (AB57737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 743 AA:
SQ

```

Query Match 15.2%; Score 64; DB 22; Length 743;  
Best Local Similarity 31.4%; Pred. No. 25;  
Matches 16; Conservative 15; Mismatches 14; Indels 6; Gaps 2;

```

QY 1 MTTRERLP-----RNGYSIAAARKLGSESTVKKRM-TSEPRFEFVARVA 45
DB 16 lprdkhnlqgrlhgdgskasvavrdlgvpestlrgwckncklrfmsrsga 66

```

Search completed: October 4, 2002, 15:34:41  
Job time: 94 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:35:15 ; Search time 28.21 Seconds  
(without alignments)  
289.528 Million cell updates/sec

Title: US-09-720-583a-3  
Perfect score: 420  
Sequence: 1 MTTRELRPRNGYSIAAANK.....EVGSVGVYHALNKRRTDA 85

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215	51.2	105	2 S32702	hypothetical prote
2	82.5	19.6	198	2 AC2478	transposase alr700
3	82	19.5	119	2 JS0053	hypothetical 13.7k
4	78	18.6	150	2 D90190	hypothetical prote
5	75.5	18.0	318	2 T35457	probable insertion
6	74.5	17.7	217	2 T31220	transposase homolo
7	74.5	17.7	217	2 A13429	transposase BME14
8	74.5	17.7	674	2 SS4178	low affinity penic
9	74	17.6	139	2 H75179	hypothetical prote
10	72.5	17.3	643	2 B75055	hypothetical prote
11	72.5	17.3	673	2 SS4175	low affinity penic
12	72.5	17.3	673	2 SS4182	low affinity penic
13	72.5	17.3	678	2 SS4176	low affinity penic
14	72.5	17.3	678	2 SS4193	low affinity penic
15	72	17.1	88	2 T43610	probable IS1617 tr
16	72	17.1	113	2 T00224	hypothetical 12.7k
17	72	17.1	125	2 F87326	transposase
18	72	17.1	131	2 T36638	probable substrate
19	72	17.1	201	2 S38652	resolvasase - Klebsi
20	72	17.1	297	2 AF3651	transcription regu
21	71	16.9	151	2 H95384	probable transcrip
22	71	16.9	297	2 E95400	probable LysK-fam1
23	71	16.9	427	1 RRYC65	RNA-binding respon
24	70	16.7	211	2 F75593	DNA-directed DNA p
25	70	16.7	297	2 C98259	hypothetical prote
26	70	16.7	297	2 AF3025	transcription regu
27	70	16.7	340	2 T43559	probable transposa
28	70	16.7	340	2 AB0096	transposase for in
29	70	16.7	340	2 AB0198	transposase for in

30	70	16.7	340	2 AB0395	transposase for in
31	70	16.7	340	2 AB0488	transposase for in
32	70	16.7	340	2 AC0031	transposase for in
33	70	16.7	340	2 AC0247	transposase for in
34	70	16.7	340	2 AC0342	transposase for in
35	70	16.7	340	2 AD0070	transposase for in
36	70	16.7	340	2 AD0124	transposase for in
37	70	16.7	340	2 AD0139	transposase for in
38	70	16.7	340	2 AD0185	transposase for in
39	70	16.7	340	2 AD0206	transposase for in
40	70	16.7	340	2 AD0457	transposase for in
41	70	16.7	340	2 AD0457	transposase for in
42	70	16.7	340	2 AE0113	transposase for in
43	70	16.7	340	2 AE0163	transposase for in
44	70	16.7	340	2 AE0254	transposase for in
45	70	16.7	340	2 AE0292	transposase for in

## ALIGNMENTS

RESULT 1  
S32702  
hypothetical protein 4 - Corynebacterium glutamicum  
C:Species: Corynebacterium glutamicum  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 02-Jun-2000  
C:Accession: S32702; S32200  
R:Shi, S.L.; Wang, Z.X.; Deng, Y.; Zhu, H.; He, B.; Zheng, Z.X.  
submitted to the EMBL Data Library, March 1993  
A:Description: Complete nucleotide sequence of a plasmid pX210142 from Corynebacteriu  
A:Reference number: S32699  
A:Accession: S32702  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-105 <SH1>  
A:Cross-references: EMBL:X72691  
C:Genetics:  
A:Start codon: GTC  
C:Superfamily: Corynebacterium glutamicum hypothetical protein 4

Query Match 51.2%; Score 215; DB 2; Length 105;  
Best Local Similarity 53.3%; Pred. No. 7.8e-15;  
Matches 40; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY 1 MTTRELRPRNGYSIAAANKIGVSESTYKRWTSEREFVARVARRHRIELRSEGSM 60  
DB 19 MTKRTRIPRNGKTIREVAGTGLSTATIERWTSAPREDYLAQANEKRVVQELRAKGLSM 78  
QY 61 RAIAAEVGSVGVYH 75  
DB 79 RAIAAEICGSVGVYH 93

RESULT 2  
AC2478  
transposase alr7003 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120alpha  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AC2478  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2478  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <KUR>  
A:Cross-references: GB:BA000020; PIDN:BA078087.1; PID:917135541; GSPDB:GN00180  
C:Experimental source: strain PCC 7120  
C:Genetics:





[illegible]

A:Cross-references: EMBL:X84862; NID:g790432; PIDN:CAB59289.1; PID:g790433  
C:Superfamily: penicillin-binding protein 2B

Query Match 17.7%; Score 74.5; DB 2; Length 674;  
Best Local Similarity 25.3%; Pred. No. 8.8;  
Matches 19; Conservative 17; Mismatches 24; Indels 15; Gaps 3;

OY 8 PRNGYSIAAA-----KKL--GVSESTVKKWTSEPRFEVAVARHARRIRLEKSGQS 59  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 366 PRTGDLALASPSPYDPKMTNGISQEDYKAYEEDPEQPSFRTGYA-----PGST 418  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 60 MRATAEVGSVGTG 74  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 419 FKMTTAAIGLDNGTT 433  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 9  
H75179  
hypothetical protein PAB2413 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C>Date: 20-Aug-1999 #sequence.revision 20-Aug-1999 #text.change 28-Jul-2000  
C:Accession: H75179  
R:Anonymous; Genoscope submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s  
A:Reference number: A75001  
A:Status: Preliminary  
A:Accession: H75019  
A:Molecule type: DNA  
A:Residues: 1-139 <KAM>  
A:Cross-references: GB:A248284; GB:AL096836; NID:g5457730; PIDN:CAB49519.1; PID:g54545  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB2413  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB2413

Query Match 17.6%; Score 74; DB 2; Length 139;  
Best Local Similarity 23.3%; Pred. No. 2;  
Matches 17; Conservative 17; Mismatches 32; Indels 4; Gaps 1;

OY 8 PRNGSIAAAAKKLGVESTVKRWTSERFEVAVARHARRIRLEKSGQSMRAIAAEV 67  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 63 PKVSRRVRALEQMGIKVIPVEKGGRPKRYSKTL----MMIEELKKGKTPREISRDL 118  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 68 GVSQTVHYALNK 80  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 119 GIPIRLTYLLRK 131  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10  
B75055  
hypothetical protein PAB1429 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C>Date: 20-Aug-1999 #sequence.revision 20-Aug-1999 #text.change 20-Aug-1999  
C:Accession: B75055  
R:Anonymous; Genoscope submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s  
A:Reference number: A75001  
A:Accession: B75055  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-643 <KAM>  
A:Cross-references: GB:A248287; GB:AL096836; NID:g5458657; PIDN:CAB50335.1; PID:e151  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1429

Query Match 17.3%; Score 72.5; DB 2; Length 643;  
Best Local Similarity 29.5%; Pred. No. 13;





**THIS PAGE BLANK (USPTO)**



```

DE RNA-directed DNA polymerase from retron Mx65 (EC 2.7.7.49) (Reverse
OS transcriptase) (Mx65-RT).
OC Myxococcus xanthus.
OC Bacteria: Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90138993; PubMed=1689062;
RA Inouye S., Herzer P.J., Inouye M.;
RT "Two independent retrons with highly diverse reverse transcriptases
RT in Myxococcus xanthus.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:942-945(1990).
CC -1- FUNCTION: PARTICIPATES IN THE SYNTHESIS OF A MSDNA (A BRANCHED RNA
CC LINKED BY A 2',5'-PHOSPHODIESTER LINKAGE TO A SINGLE-STRANDED DNA)
CC OF UNKNOWN FUNCTION. THE RETRON TRANSCRIPT SERVES AS PRIMER AND
CC TEMPLATE TO THE REACTION, AND CODES FOR THE RT.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- MISCELLANEOUS: M.XANTHUS CONTAINS TWO INDEPENDENT AND UNLINKED
CC RETRONS: Mx65 AND Mx162.
CC -1- MISCELLANEOUS: RETRONS MAY BE THE ANCESTORS OF RETROVIRUS.
CC -1- SIMILARITY: TO OTHER REVERSE TRANSCRIPTASES IN BACTERIA AND
CC RETROVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M30609; AAA88323.1; -.
DR PIR, A34864; RRYC65.
DR InterPro: IPR000123; RNA_DNAPolylms.
DR InterPro: IPR000477; RVTse.
DR Pfam: PFO0078; rvt.1
DR PRINTS: PR00866; RNADNAPOLMS.
DR TransErase; RNA-directed RNA polymerase: Transposable element.
KW SEQUENCE 427 AA; 48024 MW; C96436BF6E4D0D7 CRC64;
SQ
Query Match 16.9%; Score 71; DB 1; Length 427;
Best Local Similarity 32.48; Pred. NO. 4.6;
Matches 22; Conservative 13; Mismatches 19; Indels 14; Gaps 4;
OY 29 KRW--TTSREPEFVAARHARIRE-LRSEGSQSR-----ATAEYGVSGVY- 74
DB 77 KKMWAAARLRFSTLTTRNNRLMDLLPDEQLARYGLPWRTEDYAAAGVSGVLR 136
OY 75 HYALINKNR 82
DB 137 HYSIHRPR 144
RESULT 3
ID Y013_TREPA STANDARD; PRT; 221 AA.
AC 083057;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0013.
GN TP0013.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

```

```

RA Doderon R., Gilm W., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khaila H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-386(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; AE001186; AAC65013.1; -.
DR TIGR; TP0013; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 221 AA; 23872 MW; 5E1B311377BA9DB2 CRC64;
Cc -----
Query Match 16.5%; Score 69.5; DB 1; Length 221;
Best Local Similarity 32.2%; Pred.No.3.4;
Matches 19; Conservative 7; Mismatches 20; Indels 13; Gaps 1;
Qy 11 GYSIAAAAKKGVSESTYKRWTSPEEFEEFVAVARHAREIRLSEGSOMRAIAEVGV 69
Db 161 GYALPMWAKKNKA-----RFLVHVSPFRHLSKMELRRQRQMEAYCADIQI 206
||| | ||| | | | | | | | | | | | | | | | | | | | | | | | |
ID INN2_ECOLI STANDARD; PTR; 100 AA.
AC P39212;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insn for insertion sequence element IS911B.
DN INSN2 OR B4283.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteriophila.
OX NCBI_TaxID=562;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -!- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
CC SEQUENCE IS911.
CC -!- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; U14003; AAA97179.1; -.
DR EMBL; AE000499; AAC77239.1; -.
DR EcoGene; EG40014; INSN.
DR InterPro: IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
CW Transposable element; Transposition; DNA-binding; DNA recombination;

```

```
RW Complete proteome.
SQ SEQUENCE 100 AA; 11558 MW; 36EAC0C8EA61A52 CRC64;
```

```
Query Match      16.4%; Score 69; DB 1; Length 100;
Best Local Similarity    32.1%; Pred.No. 1.7;
Matches   17; Conservative     10; Mismatches   20; Indels       6; Gaps        1.
```

OY    12 VSIAAAAKRGVSESTVKRWITSEPPEEVAR-----VAARRARIRELRSSEQ 58  
               |::|::||::||::||::||::||::||::||::||::||::||  
DB      35 YTVADAARAAMDVGISTMTWRWKQLDEROGCKTPKASPTTPEQIEIRUKLKKLO 87

RESULT 5

ID INNI\_ECOLI STANDARD; PRT; 134 AA.  
AC P75679; P71288;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update).  
GN Transposase insn for insertion sequence element IS911A.  
OS INSIN OR B0255.  
OC Escherichia coli.  
ON Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia  
NX NCBI\_TaxID=562;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RL MEDLINE=97426617; PubMed=9278503:  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
MAU B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / W3110;  
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,  
RA Yamamoto Y., Inokuchi H., Mikii T., Hatada E., Fukuda R., Ichihara S.,  
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizubuchi K.;  
RT "Systematic sequencing of the Escherichia coli genome: Analysis of the  
4.0 - 6.0 min (189,987 - 281,416bp) region."  
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
R Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspigel N., Hyman R., Kalman S., Komp C., Kurd O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oelherer P., Roberts D.,  
RA Davis R.W.;  
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.  
CC -! FUNCTION: INVOLVED IN THE TRANSDUCTION OF DNA INSERTION  
SEQUENCE IS911.  
CC -----  
CC -! SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 8.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/eslib.ch).

CC or send an email to license@isb-sib.ch).

DR EMBL; AE000133; AAC73358.1; -.  
DR EMBL; DB3536; BAA77925.1; -.  
DR EMBL; U70214; AAB08674.1; ALT INIT.  
DR Ecocore; EG40014; lnsN.  
DR InterPro; IPRO02514; Transposase\_8.  
DR Pfam; PF01527; Transposase\_8; 1.  
KW Transposable element; transposition; DNA-binding; DNA recombination; Complete proteome.  
MW COMPLETE PROTEOME. 15249 MW; BF1AD0B16CB666AF3 CRC64;  
SQ SEQUENCE 134 AA;

```
QY      12 YSIANAANKLGVESTVARKWTSEPRFEFVAR-----VAARHARIREDRSEQ 58
Db       35 YTVADASAMDVGLSTWTRWKQLADRQGKTPKASPITPQQIEIREIRKKLQ 87
        :::::|:::||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Query Match          16.2%; Score 68; DB 1; Length 134;
Best Local Similarity 32.1%; Pred. No. 2.9; Mismatches 21; Indels 6; Gaps 1;
Matches 17; Conservative 9;

RESULT      6
YEBS_SCHPO   STANDARD; PRT; 595 AA.
AC           014301;
DT           15-JUL-1998 (Rel. 36, Created)
DF           15-JUN-1998 (Rel. 36, Last sequence update)
DE           16-OCT-2001 (Rel. 40, Last annotation update)
DN           Hypothetical 65.3 kDa Trp-Asp repeats containing protein C9g1.05 in
GE           chromosome I.
OS           SPAC9G1.05.
OC           Schizosaccharomyces pombe (fission yeast).
CC           Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC           Schizosaccharomycetales; Schizosaccharomycetaceae;
OX           NCBI_TaxID=4896;
RN           [1]
RP           SEQUENCE FROM N.A.
RC           STRAIN=972;
RA           Churcher C.M., Gentile S., Barrell B.G., Rajandream M.A., Wood V.;
RL           Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC           -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
CC           -1- SIMILARITY: BELONGS TO THE AIP1 FAMILY OF WD-REPEAT PROTEINS.
CC           -----
CC           This SWISS-PROT entry is copyright. It is produced through a collaboration
CC           between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC           at the European Bioinformatics Institute. There are no restrictions on its
CC           use by non-profit institutions as long as its content is in no way
CC           modified and this statement is not removed. Usage by and for commercial
CC           entities requires a license agreement (See http://www.isb.ch/announce/
CC           or send an email to license@isb-sib.ch).
CC           -----
DR           EMBL: Z98763; CAB1489.1; -.
DR           InterPro: IPRO01680; WD40.
DR           Pfam: PF00400; WD40; 9.
DR           PRINTS: PR00320; GPROTINBRPT.
DR           SMART: SM00320; WD40; 9.
DR           PROSITE: PS00678; WD_REPEATS_1; FALSE_NEG.
DR           PROSITE: PSS0082; WD_REPEATS_2; 6.
DR           PROSITE: PSS0294; WD_REPEATS_REGION; 1.
FM           Hypothetical protein; Repeat; WD repeat.
FT           REPEAT                    56               95             WD 1.
FT           REPEAT                    100              143            WD 2.
FT           REPEAT                    144              184            WD 3.
FT           REPEAT                    187              226            WD 4.
FT           REPEAT                    229              268            WD 5.
FT           REPEAT                    313              352            WD 6.
FT           REPEAT                    356              393            WD 7.
FT           REPEAT                    433              472            WD 8.
FT           REPEAT                    477              516            WD 9.
FT           REPEAT                    520              559            WD 10.
FT           REPEAT                    564              594            WD 11.
SO           SEQUENCE 595 AA; 65266 MW; 4F9BAFAF32823ECE CNC64;
```

OY 61 RAIA-----AEVGVSVGTV 74  
 Db 116 RIIAVGEGKERYGHAFYADSGNSVGEI 142

RESULT 7  
 CADG\_RABIT STANDARD: PRT: 829 AA.

AC Q28634;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cadherin-16 precursor (kidney-specific cadherin) (Ksp-cadherin).  
 GN COH16.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND OF SEQUENCE 570-586.  
 RC STRAIN=NEW ZEALAND WHITE;  
 RX MEDLINE=95340560; PubMed=7615566;  
 RA Thomson R.B., Igarashi P., Biemesderfer D., Kim R., Abu-Alfa A.,  
 RA Soleimani M., Aronson P.S.;  
 RT "Isolation and cDNA cloning of ksp-cadherin, a novel kidney-specific  
 RT member of the cadherin multigene family.";  
 RL J. Biol. Chem. 270:17594-17601(1995).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: KIDNEY-SPECIFIC. LIMITED TO THE BASOLATERAL  
 CC MEMBRANES OF RENAL TUBULAR EPITHELIAL CELLS.  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U28945; AAC48472.1; -.  
 DR HSSP: P09803; 150H.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF000028; cadherin; 6.  
 DR SMART: SM00112; CA: 6.  
 DR PROSITE: PS00232; CADHERIN\_1; 2.  
 DR PROSITE: PS50268; CADHERIN\_2; 6.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 1 829  
 FT DOMAIN 19 786  
 FT TRANSMEM 787 807  
 FT DOMAIN 808 829  
 FT DOMAIN 25 126  
 FT DOMAIN 131 235  
 FT DOMAIN 242 336  
 FT DOMAIN 341 449  
 FT DOMAIN 455 564  
 FT DOMAIN 569 665  
 FT DOMAIN 666 786  
 FT CARBOHYD 517 517  
 FT CARBOHYD 602 602  
 FT CARBOHYD 709 709  
 FT CARBOHYD 722 722  
 FT SEQUENCE 829 AA; 88827 MM; D2DF10E6C47A43B9 CRC64;

Query Match 15.8%; Score 66.5; DB 1; Length 829;

Best Local Similarity 26.8%; Pred. No. 26;  
 Matches 22; Conservative 12; Mismatches 17; Indels 31; Gaps 4;

OY 31 WTSEPREFEVAVARARIREI-----RSEGSMAIA-----AEVGVSVG 72  
 Db 734 WV-EPREHIVPVVVSQARVWOLPVRVYVCCNTEGECMVGKMKMPKLSAVGILVG 792

OY 73 TV-----HYALNKR 82  
 Db 793 TLAIGFFLIIFTHLARK 814

RESULT 8  
 ALPB\_RABIT STANDARD: PRT: 363 AA.

AC P79226;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).  
 GN ALDOB OR ALDB.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Beierling T.Z., Amsden A.B., Penhoet E.E., Tolian D.R.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glyceraldehyde 3-phosphate + D-glyceraldehyde 3-phosphate.  
 CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- MISCELLANEOUS: IN VERTEBRATES, THREE FORMS OF THIS UBQUITOUS  
 CC GLYCOLYTIC ENZYME ARE FOUND, ALDOLASE A IN MUSCLE, ALDOLASE B IN  
 CC LIVER, & ALDOLASE C IN BRAIN.  
 CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE  
 CC FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U85645; AAB42087.1; -.  
 DR HSSP: P00883; IADO.  
 DR InterPro: IPR000741; Aldolase\_I.  
 DR Pfam: PF00274; glycolytic-enz; 1.  
 DR ProDom: PD001128; Aldolase\_I; 1.  
 DR PROSITE: PS00158; ALDOLASE\_CLASS\_I; 1.  
 KW Lyase; Schiff base; Glycolysis; Multigene family.  
 FT INIT MET 0  
 FT BINDING 55 55  
 FT BINDING 146 146  
 FT BINDING 229 229  
 FT ACT\_SITE 363 363  
 FT SEQUENCE 363 AA; 39474 MM; 0A7185A72E89F436 CRC64;

Query Match 15.7%; Score 66; DB 1; Length 363;  
 Best Local Similarity 25.6%; Pred. No. 13;  
 Matches 22; Conservative 14; Mismatches 32; Indels 18; Gaps 2;

OY 5 ERLPRNGISIAAARKLGSESTVAKR-----TSEPREF-----VARVAR 46  
 Db 20 QRIVANGKIIILADSDESVGTGRNLRQIKRYENTENRQREILFTVDNSINOSIGVILF 79  
 OY 47 HARIRELRSEGSMAIAAEVGVSVG 72



Db 80 HETLYQKDSQKLEFRNLKEKGIYVG 105

RESULT 9

ALFB\_RAT STANDARD; PRT; 363 AA.

AC P00884; P70706; 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DE 15-DEC-1998 (Rel. 37, Last annotation update)

GN Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).

OS ALDOB.

OC Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_Taxid=10116;

RP SEQUENCE FROM N.A.

RX MEDLINE=85054935; PubMed=6094564;

RA Tsutsumi K., Mori M., Dalmon M., Tanaka T.,

RT Yatsuki H., Hori K., Ishikawa K.;

RL "Nucleotide sequence of rat liver aldolase B messenger RNA.";

RT J. Biol. Chem. 259:14572-14575(1984).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY;

RX MEDLINE=85160837; PubMed=2580098;

RA Tsutsumi K., Hidaka S., Arai Y., Hori K.,

RT Ishikawa K.;

RL "Structure and genomic organization of the rat aldolase B gene.";

RT J. Mol. Biol. 181:153-160(1985).

[3]

RP SEQUENCE OF 184-363 FROM N.A.

RX MEDLINE=83213364; PubMed=6304044;

RA Tsutsumi K., Hidaka S., Miyahara H., Tsutsumi R., Tanaka T.,

RT Hori K., Ishikawa K.;

RL "Rat aldolase isozyme gene.";

RT J. Biol. Chem. 258:6537-6542(1983).

CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glyceraldehyde + D-glyceraldehyde 3-phosphate.

CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.

CC -1- SUBUNIT: HOMOTETRAMER.

CC -1- MISCELLANEOUS: IN VERTEBRATES, THREE FORMS OF THIS UBQUITOUS GLYCOLYTIC ENZYME ARE FOUND, ALDOLASE A IN MUSCLE, ALDOLASE B IN LIVER, & ALDOLASE C IN BRAIN.

CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M10149; AAA40716.1; -

CC EMBL: X02284; CAA26156.1; -

CC EMBL: X02285; CAA26156.1; JOINED.

CC EMBL: X02286; CAA26156.1; JOINED.

CC EMBL: X02287; CAA26156.1; JOINED.

CC EMBL: X02288; CAA26156.1; JOINED.

CC EMBL: X02289; CAA26156.1; JOINED.

CC EMBL: X02290; CAA26156.1; JOINED.

CC EMBL: X02291; CAA26156.1; JOINED.

CC EMBL: Y01223; CAA24533.1; -

CC PIR: A22585; ADRTB.

CC HSSP: P00883; IADO.

CC InterPro: IPR000741; Aldolase\_I.

CC Pfam: PF00274; glycolytic\_enzy; 1.

CC ProDom: PD001128; Aldolase\_I; 1.

CC PROSITE: PS00158; ALDOLASE\_CLASS\_I; 1.

KM Lyase; Schiff base; Glycolysis; Multigene family.

FT INT\_MET 0

FT BINDING 55

FT BINDING 146

FT BINDING 229

FT ACT\_SITE 363

FT CONFLICT 233

FT SEQUENCE 363 AA; 39487 MW; 8EFAF3E621E6DD9A CRC64;

SO QUERY MATCH

Best Local Similarity 25.68; Pred. No. 13;

Matches 22; Conservative 14; Mismatches 32; Indels 18; Gaps 2;

QY 5 ERLPRNGYSIAAARKIGVSESTVQRW-----TSPRREF-----VARVAAR 46

DB 20 QRIANRKGILADESGYMGRLQRIKVENTENRKRQFRELFSVDSNISQSIGVILF 79

QY 47 HARIRELRSEGQSMRAIAEYGSVG 72

DB 80 HETLYQKDSQKLEFRNLKEKGIYVG 105

RESULT 10

MOES\_LYTVA STANDARD; PRT; 572 AA.

ID MOES\_LYTVA

AC P52962;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Moesin.

OS Lytechinus variegatus (Sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinzoa;

OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;

OC Lytechinus.

OX NCBI\_Taxid=7654;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95256329; PubMed=7738093;

RA Bachman E.S., McClay D.R.;

RT "Characterization of moesin in the sea urchin Lytechinus variegatus: RT distribution to the plasma membrane following fertilization is inhibited by cytochalasin B.";

RL J. Cell Sci. 108:161-171(1995).

CC -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL STRUCTURES TO THE PLASMA MEMBRANE.

CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U14180; AAC46514.1; -

CC InterPro: IPR000299; Band\_4.1.

CC InterPro: IPR000798; Ezrin\_radixin\_moesin.

CC Pfam: PF00373; Band\_41; 1.

CC Pfam: PF00769; ERM; 1.

CC PRINTS: PR00935; BAND41.

CC SMART: SM00295; B41; 1.

CC PROSITE: PS00660; BAND\_41.1; 1.

CC PROSITE: PS00661; BAND\_41.2; 1.

CC PROSITE: PS50057; BAND\_41.3; 1.

CC Structural protein: Cytoskeleton.

FT DOMAIN 58 224 BAND 4.1-LIKE.

SO SEQUENCE 572 AA; 67579 MW; E1BA3FB8B45E764 CRC64;

Query Match 15.7%; Score 66; DB 1; Length 572;  
 Best Local Similarity 29.9%; Pred. No. 20;  
 Matches 20; Conservative 16; Mismatches 27; Indels 4; Gaps 2;

OY 4 RRLRNNGSYIAAKKGVSESTYKVR-WTSEPREEFARV---AARHARIRELSEGOS 59  
 DB 373 QQQMSEEQSVVAAEMEMQNTATLERENTAOEKEMRARVDALEKARLQSRRESLK 432

OY 60 MRAIAE 66  
 DB 433 ESAEFAE 439

RESULT 11  
 YHJB\_ECOLI STANDARD; PRT; 200 AA.  
 AC P37640;  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical transcriptional regulator yhjb.  
 GN YHJB OR B3520.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT Nucleic Acids Res. 22:2576-2586(1994).  
 CC -1 SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: U00039; AAB18496.1; -  
 DR EMBL: AE000428; AAC76545.1; -  
 DR HSSP: P10957; IRLN.  
 DR Ecogene; EG12246; yhjB.  
 DR InterPro: IPR000792; HTH\_LuxR.  
 DR Pfam; PF00196; Gere1.1.  
 DR PRINTS; PR00038; HTHLUXR.  
 DR SMART; SM00421; HTH\_LUXR.1.  
 DR PROSITE; PS00622; HTH\_LUXR\_FAMILY.1.  
 DR Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Complete proteome.  
 KM DNA\_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).  
 SO SEQUENCE 200 AA; 22604 MW; D238AB59FE6ED109 CRC64;

Query Match 15.5%; Score 65; DB 1; Length 200;  
 Best Local Similarity 27.7%; Pred. No. 8.7;  
 Matches 18; Conservative 12; Mismatches 25; Indels 10; Gaps 2;

OY 31 WTSEPREEF--VARVAARARIRRELRSQSMRAIAEYGVGVYHVALN-----K 80  
 DB 128 WRTTPEKDKDKLSAROREILTYMLANGSNKEIGRAINISTGTVKAHLESLYRRLVYK 187

OY 81 NRTDA 85  
 DB 188 NRTQA 192

RESULT 12  
 ATCU\_RHIME STANDARD; PRT; 827 AA.  
 ID ATCU\_RHIME  
 AC Q9X5X3;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Copper-transporting ATPase (EC 3.6.3.4).  
 GN ACTP.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Sinorhizobium.  
 CC NCBI\_TaxID=362;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WM5419;  
 RA Reeve W.G., Tlwall R.P., Kale N.B., Dilworth M.J., Glenn A.R.;

RT "The role of copper and P-type ATPase in the acid-tolerance of  
 RT Rhizobium leguminosarum bv viciae and Sinorhizobium meliloti.";  
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1 FUNCTION: INVOLVED IN COPPER TRANSPORT (BY SIMILARITY).  
 CC -1 CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1 SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (EI-E2 ATPASES). SUBFAMILY 1B.  
 CC -1 SIMILARITY: CONTAINS 2 HMA DOMAINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AF129004; AAD27639.1; -  
 DR HSSP: Q04656; IAW0.  
 DR InterPro: IPR001366; Cad\_ATPase.  
 DR InterPro: IPR000579; Cat\_P\_ATPase.  
 DR InterPro: IPR001757; EI-E2\_ATPase.  
 DR InterPro: IPR001802; HG\_scatvenger.  
 DR InterPro: IPR001934; HMA.  
 DR InterPro: IPR001454; Hydrolyase.  
 DR Pfam; PF00122; EI-E2\_ATPase; 1.  
 DR Pfam; PF00403; HMA; 2.  
 DR Pfam; PF00702; Hydrolyase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00940; CATATPASEA.  
 DR PRINTS; PR00941; CDATPASE.  
 DR PRINTS; PR00946; HGSCAVENGER.  
 DR PROSITE; PS00154; ATPASE\_EI-E2; 1.  
 DR PROSITE; PS01047; HMA\_1; 2.  
 DR PROSITE; PS00846; HMA\_2; 2.  
 DR Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Metal binding; Copper; Repeat.  
 KM Metal binding; Copper; Repeat.

FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 210 230 POTENTIAL.  
 FT TRANSMEM 248 268 POTENTIAL.  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT TRANSMEM 430 450 POTENTIAL.  
 FT TRANSMEM 458 478 POTENTIAL.  
 FT TRANSMEM 541 561 POTENTIAL.  
 FT TRANSMEM 569 589 POTENTIAL.  
 FT TRANSMEM 727 747 POTENTIAL.  
 FT TRANSMEM 773 793 POTENTIAL.  
 FT TRANSMEM 795 815 POTENTIAL.  
 FT DOMAIN 16 81 HMA 1.  
 FT DOMAIN 83 149 HMA 2.  
 FT METAL 26 26 COPPER (POTENTIAL).  
 FT METAL 29 29 COPPER (POTENTIAL).  
 FT METAL 93 93 COPPER (POTENTIAL).  
 FT METAL 96 96 COPPER (POTENTIAL).  
 FT MOD\_RES 515 515 PHOSPHORYLATION (BY SIMILARITY).

DT 01-MAR-2002 (Rel. 41, last sequence update)  
DT 01-MAR-2002 (Rel. 41, last annotation update)  
DE Copper-transporting ATPase 2 (EC 3.6.3.4).

GN ACTP2 OR ATC2 OR RB1018 OR SMB21578.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Plasmid pSymB (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1021;  
RX MEDLINE-21396508; PubMed-11481431;  
RA Flann T.M., Weidner S., Wong K., Bunrmester J., Chain P.,  
RA Vorhoelster F.J., Hernandez-Lucas I., Becker A., Cowie A., Guzy J.,  
RA Golding B., Puchler A.,  
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-  
fixing endosymbiont Sinorhizobium meliloti."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
CC -1- FUNCTION: INVOLVED IN COPPER TRANSPORT (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
CC -1- SIMILARITY: SUBFAMILY IB.  
CC -1- SIMILARITY: CONTAINS 2 HMA DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AL603645; CAC9418.1; -;  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
DR PROSITE; PS01047; HMA\_1; 2.  
DR PROSITE; PS50846; HMA\_2; 2.  
KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
KW Metal-binding; Copper; Repeat; Plasmid; Complete proteome.  
FT TRANSMEM 174 194 POTENTIAL.  
FT TRANSMEM 210 230 POTENTIAL.  
FT TRANSMEM 246 266 POTENTIAL.  
FT TRANSMEM 271 291 POTENTIAL.  
FT TRANSMEM 430 450 POTENTIAL.  
FT TRANSMEM 458 478 POTENTIAL.  
FT TRANSMEM 771 793 POTENTIAL.  
FT TRANSMEM 797 819 POTENTIAL.  
FT DOMAIN 16 81 HMA 1.  
FT DOMAIN 83 149 HMA 2.  
FT METAL 26 26 COPPER (POTENTIAL).  
FT METAL 29 29 COPPER (POTENTIAL).  
FT METAL 93 93 COPPER (POTENTIAL).  
FT METAL 96 96 COPPER (POTENTIAL).  
FT MOD\_RES 515 515 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 714 714 MAGNESIUM (BY SIMILARITY).  
FT METAL 718 718 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 827 AA; 85861 MW; A3D8DFDD1315FCB CRC64;  
  
Query Match 15.2%; Score 64; DB 1; Length 827;  
Best Local Similarity 29.9%; Pred. No. 47;  
Matches 23; Conservative 11; Mismatches 17; Indels 26; Gaps 3;  
  
QY 10 NGYSTAAAKKLGVSSTVKTSTPREPVARV--AARHARIRLRSGQSMRAIA--- 64  
Db 670 NRRFAEAIARKLGI-----DEVAEVLPEKGVKAIKRLRGGRSVAVTIGDGI 716  
QY 65 -----AEVCVSYGT 73  
Db 717 NDAPALAEADVGIAGVT 733

---

**THIS PAGE BLANK (USPTO)**